GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:04 ; Search time 47.36 Seconds
(without alignments)
292.164 Million cell updates/sec

Title: US-08-981-087A-2
Perfect score: 144
Sequence: 1 SYTNDKILILYFNKLYKKIK.....LNYNKIIWTLQDTAGNNQKL 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144	100.0	366	2	S48110	neurotoxin type F
2	26	18.1	369	2	S48109	neurotoxin type F
3	26	18.1	1274	2	I40813	neurotoxin type F
4	15	10.4	1268	2	S33411	botulinum neurotox
5	11	7.6	367	2	S48106	neurotoxin type E
6	11	7.6	1251	2	JH0256	botulinum neurotox
7	11	7.6	1252	2	S21178	botulinum neurotox
8	9	6.2	1296	2	I40645	botulinum neurotox
9	8	5.6	1072	2	A86827	hypothetical prote
10	8	5.6	1291	1	A48940	bontoxilysin (EC 3
11	8	5.6	1291	2	I40631	non-proteolytic bo
12	8	5.6	1296	1	BTCLAB	bontoxilysin (EC 3
13	7	4.9	177	2	B70190	conserved hypothet
14	7	4.9	209	2	G90445	hypothetical prote
15	7	4.9	213	2	AC2379	hypothetical prote
16	7	4.9	241	2	B90474	conserved hypothet
17	7	4.9	244	2	T28307	ORF MSV146 hypothe
18	7	4.9	261	2	G84453	probable GDSL-moti
19	7	4.9	330	2	E89791	hypothetical prote
20	7	4.9	405	2	AC3644	alpha-methylacyl-C
21	7	4.9	435	2	AI2082	hypothetical prote
22	7	4.9	449	2	F71664	UDP-N-acetylmuramo
23	7	4.9	484	2	T25731	hypothetical prote
24	7	4.9	501	1	VGBEMA	glycoprotein A pre
25	7	4.9	501	1	VGBEMB	glycoprotein A pre
26	7	4.9	501	1	A60005	glycoprotein A pre
27	7	4.9	505	1	VGBEMH	glycoprotein gp57-
28	7	4.9	511	2	S53061	hypothetical prote
29	7	4.9	737	2	C70132	hypothetical prote
30	7	4.9	1010	2	T13167	Lola-like protein
31	7	4.9	1116	2	T16112	hypothetical prote
32	7	4.9	1711	2	C71625	variant-specific s
33	7	4.9	1900	2	AG2391	serine/threonine k
34	7	4.9	2292	1	GNNYED	genome polyprotein
35	7	4.9	2292	1	GNNYEB	genome polyprotein
36	7	4.9	2292	2	S55401	capsid polyprotein
37	6	4.2	30	2	PD0006	cysteine synthase
38	6	4.2	56	2	S66323	protein kinase AK1
39	6	4.2	57	2	S66314	protein kinase AK1
40	6	4.2	73	2	T13199	hypothetical prote
41	6	4.2	79	2	I38991	tumorigenic conver
42	6	4.2	97	2	A86848	hypothetical prote
43	6	4.2	99	2	G90113	Rbp1 protein [impo
44	6	4.2	101	2	S72281	hypothetical prote
45	6	4.2	112	2	C64498	hypothetical prote

ALIGNMENTS

RESULT 1
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 100.0%; Score 144; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 4e-141;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     214 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 273

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     274 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 333

Qy     121 WKISLNYNKIIWTLQDTAGNNQKL 144
          ||||||||||||||||||||||||
Db     334 WKISLNYNKIIWTLQDTAGNNQKL 357
```

RESULT 2
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993

C;Superfamily: tetanus toxin

Query Match 18.1%; Score 26; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 5.6e-19;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 SILDMRYENNKFIDISGYGSNISING 48
||||||||||||||||||||||||||
Db 236 SILDMRYENNKFIDISGYGSNISING 261

RESULT 3
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson, D.E
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 18.1%; Score 26; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 SILDMRYENNKFIDISGYGSNISING 48
||||||||||||||||||||||||||
Db 869 SILDMRYENNKFIDISGYGSNISING 894

RESULT 4
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 10.4%; Score 15; DB 2; Length 1268;
Best Local Similarity 100.0%; Pred. No. 3.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 YIYSTNRNQFGIYSS 65
|||||||||||||||
Db 889 YIYSTNRNQFGIYSS 903

RESULT 5
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 7.6%; Score 11; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 NFSISFWVRIP 96
||||||||||||
Db 299 NFSISFWVRIP 309

RESULT 6
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum t
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa,
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E t
A;Reference number: S16145; MUID:91237316
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter relea
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 7.6%; Score 11; DB 2; Length 1251;
Best Local Similarity 100.0%; Pred. No. 0.0052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 NFSISFWVRIP 96

```
          ||||||||||
Db    914 NFSISFWVRIP 924


RESULT   7
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted


  Query Match             7.6%;  Score 11;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 0.0052;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||
Db    914 NFSISFWVRIP 924


RESULT   8
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type
A;Reference number: I40645; MUID:94143603
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin


  Query Match             6.2%;  Score 9;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.63;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     103 NLNNEYTII 111
           |||||||||
Db    957 NLNNEYTII 965


RESULT   9
A86827
hypothetical protein yqfG [imported] - Lactococcus lactis subsp. lactis (strain IL140
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: A86827
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: A86827
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1072 <STO>
A;Cross-references: GB:AE005176; PID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqfG


  Query Match             5.6%;  Score 8;  DB 2;  Length 1072;
  Best Local Similarity   100.0%;  Pred. No. 5.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      64 SSKPSEVN 71
           ||||||||
Db    712 SSKPSEVN 719


RESULT  10
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the
A;Reference number: A48940; MUID:92384550
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
```

A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta, B.R
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolyt
A;Reference number: S27125; MUID:93063293
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 5.6%; Score 8; DB 1; Length 1291;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 NNSGWKIS 124
||||||||
Db 958 NNSGWKIS 965

RESULT 11
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botul
A;Reference number: I40631; MUID:94122659
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalyti
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted

Query Match 5.6%; Score 8; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 117 NNSGWKIS 124
||||||||
Db 958 NNSGWKIS 965

RESULT 12
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with oth

A;Reference number: A35294; MUID:90264400
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson, T.
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotoxin
A;Reference number: S09492; MUID:90235864
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, an
A;Reference number: A53884; MUID:94000342
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 5.6%; Score 8; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 LNNEYTII 111
||||||||
Db 958 LNNEYTII 965

RESULT 13
B70190
conserved hypothetical protein BB0723 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 26-Aug-1999
C;Accession: B70190
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: B70190
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-177 <KLE>
A;Cross-references: GB:AE000783; TIGR:BB0723
A;Experimental source: strain B31
C;Superfamily: conserved hypothetical protein MJ0240

Query Match 4.9%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 KLYKKIK 20
|||||||
Db 98 KLYKKIK 104

RESULT 14
G90445
hypothetical protein SSO2711 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: G90445
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: G90445
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <KUR>
A;Cross-references: GB:AE006641; NID:g13816031; PIDN:AAK42822.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2711

Query Match 4.9%; Score 7; DB 2; Length 209;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 NKFIDIS 38
|||||||
Db 53 NKFIDIS 59

RESULT 15
AC2379

hypothetical protein alr4587 [imported] - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C;Accession: AC2379
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AC2379
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-213 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB76286.1; PID:g17133723; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr4587
C;Superfamily: biopolymer transport protein exbB-1

Query Match 4.9%; Score 7; DB 2; Length 213;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      89 ISFWVRI 95
           |||||||
Db      32 ISFWVRI 38
```

Search completed: August 15, 2002, 11:14:06
Job time: 258 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:09:48 ; Search time 47.36 Seconds
(without alignments)
874.462 Million cell updates/sec

Title: US-08-981-087A-1
Perfect score: 431
Sequence: 1 SYTNDKILILYFNKLYKKIK.....TSSNGCFWSFISKEHGWQEN 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	153	35.5	366	2	S48110	neurotoxin type F
2	27	6.3	1274	2	I40813	neurotoxin type F
3	26	6.0	369	2	S48109	neurotoxin type F
4	22	5.1	1268	2	S33411	botulinum neurotox
5	15	3.5	1251	2	JH0256	botulinum neurotox
6	15	3.5	1252	2	S21178	botulinum neurotox
7	11	2.6	367	2	S48106	neurotoxin type E
8	11	2.6	1296	1	BTCLAB	bontoxilysin (EC 3
9	11	2.6	1296	2	I40645	botulinum neurotox
10	10	2.3	458	2	T02571	probable myrosinas
11	8	1.9	540	2	B87350	hypothetical prote
12	8	1.9	1072	2	A86827	hypothetical prote
13	8	1.9	1291	1	A48940	bontoxilysin (EC 3
14	8	1.9	1291	2	I40631	non-proteolytic bo
15	8	1.9	1297	2	S39791	neurotoxin - Clost
16	7	1.6	67	2	AI1594	hypothetical prote
17	7	1.6	156	2	S20990	myosin regulatory
18	7	1.6	177	2	B70190	conserved hypothet
19	7	1.6	200	2	A81295	probable membrane
20	7	1.6	202	2	S20992	myosin regulatory
21	7	1.6	204	2	S22715	myosin regulatory
22	7	1.6	209	2	G90445	hypothetical prote
23	7	1.6	213	2	AC2379	hypothetical prote
24	7	1.6	226	2	AI2668	hypothetical prote
25	7	1.6	231	2	S28494	dtxA protein - Clo
26	7	1.6	241	2	B90474	conserved hypothet
27	7	1.6	242	2	T27999	hypothetical prote
28	7	1.6	243	2	A91044	hypothetical prote
29	7	1.6	243	2	D85888	hypothetical prote
30	7	1.6	244	2	T28307	ORF MSV146 hypothe
31	7	1.6	259	2	AF1544	RNA polymerase sig
32	7	1.6	259	2	AG1186	RNA polymerase sig
33	7	1.6	261	2	G84453	probable GDSL-moti
34	7	1.6	264	2	G70195	pyridoxal kinase (
35	7	1.6	265	2	AG0814	cob(I)alamin adeno
36	7	1.6	267	2	G97450	ABC transporter, A
37	7	1.6	267	2	F81029	type II restrictio
38	7	1.6	267	2	B65021	cob(I)alamin adeno
39	7	1.6	280	2	T47572	Machado-Joseph dis
40	7	1.6	307	2	C83188	probable ATP-bindi
41	7	1.6	330	2	E89791	hypothetical prote
42	7	1.6	347	2	T19989	hypothetical prote
43	7	1.6	358	2	S07594	hypothetical prote
44	7	1.6	375	2	A83636	hypothetical prote
45	7	1.6	379	2	B69344	hypothetical prote

ALIGNMENTS

RESULT 1
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 35.5%; Score 153; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 6.6e-149;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     214 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 273

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     274 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 333

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMIS 153
           |||||||||||||||||||||||||||||||||
Db     334 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMIS 366
```

RESULT 2
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson,
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.

J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 6.3%; Score 27; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 157 YINKWIFVTITNNRLGNSRIYINGNLI 183
||||||||||||||||||||||||||||
Db 1006 YINKWIFVTITNNRLGNSRIYINGNLI 1032

RESULT 3
S48109
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 16-Jul-1999
C;Accession: S48109
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48109
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <CAM>
A;Cross-references: EMBL:X70820; NID:g407790; PIDN:CAA50151.1; PID:g407791
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin

Query Match 6.0%; Score 26; DB 2; Length 369;
Best Local Similarity 100.0%; Pred. No. 2.2e-18;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 SILDMRYENNKFIDISGYGSNISING 48
||||||||||||||||||||||||||
Db 236 SILDMRYENNKFIDISGYGSNISING 261

RESULT 4
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P.T.
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neurotoxin
A;Reference number: S33411; MUID:93252228
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 5.1%; Score 22; DB 2; Length 1268;
Best Local Similarity 100.0%; Pred. No. 9.2e-14;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 198 DNILFKIVGCNDTRYVGIRYFK 219
||||||||||||||||||||||
Db 1036 DNILFKIVGCNDTRYVGIRYFK 1057

RESULT 5
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum t
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa,
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E t
A;Reference number: S16145; MUID:91237316
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter relea
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 2; Length 1251;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 ISDYINKWIFVTITN 168
|||||||||||||||
Db 983 ISDYINKWIFVTITN 997

RESULT 6
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neuroto
A;Reference number: S21178; MUID:92174922
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.

Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted

Query Match 3.5%; Score 15; DB 2; Length 1252;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 ISDYINKWIFVTITN 168
||||||||||||||||
Db 983 ISDYINKWIFVTITN 997

RESULT 7
S48106
neurotoxin type E - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48106
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48106
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-367 <CAM>
A;Cross-references: EMBL:X70818; NID:g407784; PIDN:CAA50149.1; PID:g407785
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 2.6%; Score 11; DB 2; Length 367;
Best Local Similarity 100.0%; Pred. No. 0.0058;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 NFSISFWVRIP 96
||||||||||||
Db 299 NFSISFWVRIP 309

RESULT 8
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with oth
A;Reference number: A35294; MUID:90264400
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson,
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neurotc
A;Reference number: S09492; MUID:90235864
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin compone
A;Reference number: S67988; MUID:96096783
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-t
A;Reference number: A33401; MUID:89350959
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42,
A;Reference number: A53884; MUID:94000342
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and a
A;Reference number: A60025; MUID:91120847
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic sy

C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 2.6%; Score 11; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 WIFVTITNNRL 171
||||||||||||
Db 1014 WIFVTITNNRL 1024

RESULT 9
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A
A;Reference number: I40645; MUID:94143603
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 2.6%; Score 11; DB 2; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.019;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 161 WIFVTITNNRL 171
||||||||||||
Db 1014 WIFVTITNNRL 1024

RESULT 10
T02571
probable myrosinase-binding protein [imported] - Arabidopsis thaliana
N;Alternate names: hypothetical protein T16B24.5
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C;Accession: T02571; G84815
R;Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
submitted to the EMBL Data Library, August 1998
A;Description: Arabidopsis thaliana chromosome II BAC T16B24 genomic sequence.
A;Reference number: Z14679
A;Accession: T02571
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-458 <ROU>
A;Cross-references: EMBL:AC004697; NID:g3402671; PID:g3402676
A;Experimental source: cultivar Columbia
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84815
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-458 <STO>
A;Cross-references: GB:AE002093; NID:g3402676; PIDN:AAC28979.1; GSPDB:GN00139
C;Genetics:
A;Gene: T16B24.5; At2g39310
A;Map position: 2
A;Introns: 67/3; 221/3; 374/3

Query Match 2.3%; Score 10; DB 2; Length 458;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 285 YQKPNIFSNT 294
||||||||||
Db 238 YQKPNIFSNT 247

RESULT 11
B87350
hypothetical protein CC0813 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: B87350
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87350
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-540 <STO>
A;Cross-references: GB:AE005673; NID:g13422062; PIDN:AAK22798.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0813

Query Match 1.9%; Score 8; DB 2; Length 540;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 238 PDPSILKD 245
||||||||
Db 71 PDPSILKD 78

RESULT 12
A86827
hypothetical protein yqfG [imported] - Lactococcus lactis subsp. lactis (strain IL140
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C;Accession: A86827
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: A86827
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1072 <STO>
A;Cross-references: GB:AE005176; PID:g12724625; PIDN:AAK05715.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: yqfG

Query Match 1.9%; Score 8; DB 2; Length 1072;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      64 SSKPSEVN 71
           ||||||||
Db     712 SSKPSEVN 719


RESULT  13
A48940
bontoxilysin (EC 3.4.24.69) B precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type B (BoNT/B)
C;Species: Clostridium botulinum
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A48940; S48105; S21575; A42871; S07155; S08562; S07128; S08573; S08574
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Brehm, J.K.; Atkinson, T.; Minton, N.P.
Appl. Environ. Microbiol. 58, 2345-2354, 1992
A;Title: Molecular cloning of the Clostridium botulinum structural gene encoding the typ
A;Reference number: A48940; MUID:92384550
A;Accession: A48940
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1291 <WHE>
A;Cross-references: GB:M81186; NID:g144734; PIDN:AAA23211.1; PID:g144735
A;Experimental source: type B, Danish
A;Note: sequence extracted from NCBI backbone (NCBIN:112080, NCBIP:112081); this publica
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48105
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-994 <CAM>
A;Cross-references: EMBL:X70817; NID:g407782; PIDN:CAA50148.1; PID:g407783
A;Experimental source: proteolytic type B, strain NCTC 7273
R;Szabo, E.A.; Pemberton, J.M.; Desmarchelier, P.M.
submitted to the EMBL Data Library, April 1992
A;Description: Partial amino acid sequence of botulinum neurotoxin type B and comparisic
A;Reference number: S21575
A;Accession: S21575
A;Molecule type: DNA
A;Residues: 36-217,'G',219-224,'S',226-246 <SZA>
A;Cross-references: EMBL:Z11934; NID:g40383; PIDN:CAA77991.1; PID:g40384
R;Kurazono, H.; Mochida, S.; Binz, T.; Eisel, U.; Quanz, M.; Grebenstein, O.; Wernars, K
J. Biol. Chem. 267, 14721-14729, 1992
A;Title: Minimal essential domains specifying toxicity of the light chains of tetanus to
A;Reference number: A42871; MUID:92340509
A;Accession: A42871
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-313,'S',315-451 <KUR>
A;Experimental source: strain Okra
A;Note: sequence extracted from NCBI backbone (NCBIP:109365)
R;DasGupta, B.R.; Datta, A.
Biochimie 70, 811-817, 1988
A;Title: Botulinum neurotoxin type B (strain 657): partial sequence and similarity with
A;Reference number: S07155; MUID:89000987
A;Accession: S07155
A;Molecule type: protein
A;Residues: 2-29,'M',31-45 <DAS>
A;Accession: S08562
A;Molecule type: protein
A;Residues: 442-463,'R',465-467 <DA2>
R;Schmidt, J.J.; Sathyamoorthy, V.; DasGupta, B.R.
Arch. Biochem. Biophys. 238, 544-548, 1985
A;Title: Partial amino acid sequences of botulinum neurotoxins types B and E.
A;Reference number: S07128; MUID:85197963
A;Accession: S07128
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-16 <SCH1>
A;Accession: S08573
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-17 <SCH2>
A;Accession: S08574
A;Status: preliminary
A;Molecule type: protein
A;Residues: 442-459 <SCH3>
R;Schiavo, G.; Benfenati, F.; Poulain, B.; Rossetto, O.; de Laureto, P.P.; DasGupta,
Nature 359, 832-835, 1992
A;Title: Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteo
A;Reference number: S27125; MUID:93063293
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic sy
C;Genetics:
A;Gene: bont/b
C;Function:
A;Description: catalyzes hydrolysis of a Gln-Phe peptide bond in synaptobrevin 2
C;Superfamily: tetanus toxin
C;Keywords: hydrolase; metalloproteinase; neurotoxin; transmembrane protein; zinc
F;2-441/Product: bontoxilysin B light chain #status experimental <LGHT>
F;442-1291/Product: bontoxilysin B heavy chain #status experimental <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted


  Query Match             1.9%;  Score 8;  DB 1;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     958 NNSGWKIS 965


RESULT  14
I40631
non-proteolytic botulinum neurotoxin type B precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40631; S48103; S48104; S36015
R;Hutson, R.A.; Collins, M.D.; East, A.K.; Thompson, D.E.
Curr. Microbiol. 28, 101-110, 1994
A;Title: Nucleotide sequence of the gene coding for non-proteolytic Clostridium botul
A;Reference number: I40631; MUID:94122659
A;Accession: I40631
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1291 <RES>
A;Cross-references: EMBL:X71343; NID:g296148; PIDN:CAA50482.1; PID:g296149
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48103
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 634-761,'E',763-841,'M',843,'T',845,'N',847-994 <CAM1>
A;Cross-references: EMBL:X70814; NID:g407778; PIDN:CAA50145.1; PID:g407779
A;Experimental source: non-proteolytic strain 2129B (Scott)
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
A;Accession: S48104
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 634-843,'T',845,'N',847-994 <CAM2>
A;Cross-references: EMBL:X70819; NID:g407780; PIDN:CAA50150.1; PID:g407781
A;Experimental source: non-proteolytic strain Eklund 2B (Colworth 229)
C;Comment: Botulinum neurotoxin type B in these strains may posses a capable catalyti
C;Genetics:
A;Gene: bont/b
C;Superfamily: tetanus toxin
C;Keywords: metalloprotein; neurotoxin; transmembrane protein; zinc
F;2-441/Product: botulinum neurotoxin type B light chain #status predicted <LGHT>
F;442-1291/Product: botulinum neurotoxin type B heavy chain #status predicted <HVY>
F;230,234/Binding site: zinc (His) #status predicted
F;231/Active site: Glu #status predicted
```

Query Match 1.9%; Score 8; DB 2; Length 1291;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     117 NNSGWKIS 124
           ||||||||
Db     958 NNSGWKIS 965
```

RESULT 15
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 1.9%; Score 8; DB 2; Length 1297;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     154 ISDYINKW 161
           ||||||||
Db    1002 ISDYINKW 1009
```

Search completed: August 15, 2002, 11:14:04
Job time: 256 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:07:08 ; Search time 96.53 Seconds
(without alignments)
495.937 Million cell updates/sec

Title: US-08-981-087A-1
Perfect score: 431
Sequence: 1 SYTNDKILILYFNKLYKKIK.....TSSNGCFWSFISKEHGWQEN 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	431	100.0	431	18	AAW09014	Immunogenic type F
2	431	100.0	432	22	AAB04096	Botulism toxin hea
3	431	100.0	432	22	AAB04103	Botulism toxin hea
4	431	100.0	645	22	AAE07894	Modified clostridi
5	431	100.0	685	22	AAE07893	Modified clostridi
6	431	100.0	862	22	AAE07890	Modified clostridi
7	431	100.0	887	22	AAE07892	Modified clostridi
8	431	100.0	1032	22	AAE07901	C. botulinum C2 tr
9	431	100.0	1059	21	AAY93309	A manganese supero
10	431	100.0	1084	21	AAY93312	A manganese supero
11	431	100.0	1092	22	AAE07900	C. botulinum C2 tr
12	213	49.4	660	22	AAE07898	Modified clostridi
13	212	49.2	432	21	AAY77138	Synthetic botulinu
14	144	33.4	144	18	AAW09015	Immunogenic type F
15	144	33.4	144	18	AAW09016	Immunogenic type F
16	143	33.2	143	18	AAW09017	Immunogenic type F
17	27	6.3	448	19	AAW68399	Clostridium botuli
18	15	3.5	419	22	AAB04095	Botulism toxin hea
19	15	3.5	449	21	AAY77137	Synthetic botulinu
20	15	3.5	449	22	AAB04094	Botulism toxin hea
21	15	3.5	451	19	AAW68395	Clostridium botuli
22	15	3.5	452	19	AAW68396	Clostridium botuli
23	11	2.6	233	21	AAY77143	Botulinum neurotox
24	11	2.6	382	21	AAB36303	BoNT/A prototoxin
25	11	2.6	415	22	AAB04083	Botulism toxin C f
26	11	2.6	432	21	AAY77142	Native botulinum n
27	11	2.6	434	22	AAB04089	Botulism toxin hea
28	11	2.6	435	22	AAB04090	Botulism toxin hea
29	11	2.6	437	22	AAB04088	Botulism toxin hea
30	11	2.6	438	17	AAR95008	Type A neurotoxin
31	11	2.6	438	19	AAW68389	Clostridium botuli
32	11	2.6	438	21	AAY77134	Synthetic botulinu
33	11	2.6	445	19	AAW68391	Clostridium botuli
34	11	2.6	462	17	AAR95009	Type A neurotoxin
35	11	2.6	462	19	AAW68390	Clostridium botuli
36	11	2.6	837	21	AAY77140	Native botulinum n
37	11	2.6	847	22	AAB04081	Botulism toxin hea
38	11	2.6	1067	21	AAY93307	A manganese supero
39	11	2.6	1092	21	AAY93310	A manganese supero
40	11	2.6	1296	17	AAR95010	C. botulinum type
41	10	2.3	455	21	AAG16791	Arabidopsis thalia
42	10	2.3	455	21	AAG42811	Arabidopsis thalia
43	10	2.3	458	21	AAG16790	Arabidopsis thalia
44	10	2.3	458	21	AAG42810	Arabidopsis thalia
45	8	1.9	439	22	AAB04085	Botulism toxin C f

ALIGNMENTS

RESULT 1
AAW09014
ID AAW09014 standard; Protein; 431 AA.
XX
AC AAW09014;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX

6-12-95.

```
CC   A polypeptide (AAW09014) comprises the heavy chain (amino acids
CC   848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be
CC   produced using a synthetic gene (AAT48101) based on the natural
CC   gene sequence (AAT48100) for the heavy chain.  The polypeptides and
CC   its fragments (see also AAW09015-17) lack the light chain and HN
CC   epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   431 AA;


  Query Match             100.0%;  Score 431;  DB 18;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 60

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 120

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 180

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 240

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 300

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 360

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 420

Qy     421 FISKEHGWQEN 431
           |||||||||||
Db     421 fiskehgwqen 431


RESULT   2
AAB04096
ID   AAB04096 standard; Protein; 432 AA.
XX
AC   AAB04096;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54490.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Claim 3; Fig 9b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 181

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 241

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 301

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 361

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    362 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 421

Qy    421 FISKEHGWQEN 431
          ||||||||||| |
Db    422 fiskehgwqen 432


RESULT   3
AAB04103
ID   AAB04103 standard; Protein; 432 AA.
XX
AC   AAB04103;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54499.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Disclosure; Fig 18b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 0;

  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 181

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 241

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 301

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 361

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 421

Qy    421 FISKEHGWQEN 431
          ||||||||||| |
Db    422 fiskehgwqen 432


RESULT   4
AAE07894
ID   AAE07894 standard; Protein; 645 AA.
XX
AC   AAE07894;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 44; 50pp; English.
XX
```

```
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain fragment. This sequence is
CC   constructed by fusing the binding domain of botulinum neurotoxin type F
CC   (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ   Sequence   645 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 645;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     215 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 274

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     275 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 334

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     335 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 394

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     395 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 454

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     455 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 514

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     515 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 574

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     575 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 634

Qy     421 FISKEHGWQEN 431
           |||||||||||
Db     635 fiskehgwqen 645


RESULT   5
AAE07893
ID   AAE07893 standard; Protein; 685 AA.
XX
AC   AAE07893;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Influenza virus.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 43; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC   conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation peptide from influenza virus and a neuronal cell-specific
CC   binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   685 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 314

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 374

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     375 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 434

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     435 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 494

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     495 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 554

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     555 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 614
```

```
Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     615 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 674

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db     675 fiskehgwqen 685


RESULT   6
AAE07890
ID   AAE07890 standard; Protein; 862 AA.
XX
AC   AAE07890;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 40; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate.
CC   This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation domain from diphtheria neurotoxin and a neuronal cell-
CC   specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   862 AA;
```

```
  Query Match           100.0%;  Score 431;  DB 22;  Length 862;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 491

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 551

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     552 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 611

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     612 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 671

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     672 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 731

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     732 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 791

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     792 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 851

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db     852 fiskehgwqen 862


RESULT   7
AAE07892
ID   AAE07892 standard; Protein; 887 AA.
XX
AC   AAE07892;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
```

```
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 42; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate.
CC   This conjugate comprises a mitochondrial leader sequence from human
CC   Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC   linker that can be cleaved by thrombin, translocation domain from
CC   diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC   botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   887 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 887;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 516

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     517 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 576

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     577 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 636

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     637 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 696

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     697 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 756

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     757 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 816

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     817 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 876

Qy     421 FISKEHGWQEN 431
           ||||||||||| 
Db     877 fiskehgwqen 887


RESULT   8
AAE07901
ID   AAE07901 standard; Protein; 1032 AA.
XX
AC   AAE07901;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 48; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1032 AA;


  Query Match             100.0%;  Score 431;  DB 22;  Length 1032;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 661

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 721

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 781

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 841

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 901
```

```
Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    902 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 961

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    962 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1021

Qy    421 FISKEHGWQEN 431
          ||||||||||| 
Db   1022 fiskehgwqen 1032


RESULT   9
AAY93309
ID   AAY93309 standard; protein; 1059 AA.
XX
AC   AAY93309;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC   can be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a
CC   domain that translocates the SOD of the composition into the neuronal
CC   cell. After translocation, the linker is cleaved to release the SOD.
CC   The composition is useful for treating neuronal diseases caused or
CC   augmented by oxidative stress, such as ischemic stroke, trauma,
CC   Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ   Sequence   1059 AA;

  Query Match           100.0%;  Score 431;  DB 21;  Length 1059;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 688

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 748

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    749 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 808

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    809 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 868

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    869 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 928

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    929 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 988

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    989 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1048

Qy    421 FISKEHGWQEN 431
          ||||||||||| 
Db   1049 fiskehgwqen 1059


RESULT  10
AAY93312
ID   AAY93312 standard; protein; 1084 AA.
XX
AC   AAY93312;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
```

```
PS   Disclosure; Page 57-60; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide
CC   dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC   that can be cleaved by thrombin, and a heavy chain derived from
CC   botulinum neurotoxin serotype F. The specification describes a
CC   composition for delivery of SOD to neuronal cells. The composition
CC   comprises SOD linked, by a cleavable linker, to a neuronal cell
CC   targeting component (NCTC). This component has a domain that binds
CC   to a neuronal cell and a domain that translocates the SOD of the
CC   composition into the neuronal cell. After translocation, the linker
CC   is cleaved to release the SOD. The composition is useful for treating
CC   neuronal diseases caused or augmented by oxidative stress, such as
CC   ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases.
XX
SQ   Sequence   1084 AA;


  Query Match           100.0%;  Score 431;  DB 21;  Length 1084;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     654 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 713

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 773

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     774 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 833

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     834 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 893

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     894 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 953

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 1013

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1073

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db    1074 fiskehgwqen 1084


RESULT  11
AAE07900
ID   AAE07900 standard; Protein; 1092 AA.
XX
AC   AAE07900;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 47; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1092 AA;


  Query Match           100.0%;  Score 431;  DB 22;  Length 1092;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 721

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 781

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 841

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 901

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 961

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     962 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 1021

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1022 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1081
```

```
Qy    421 FISKEHGWQEN 431
          ||||||||||||
Db   1082 fiskehgwqen 1092


RESULT  12
AAE07898
ID   AAE07898 standard; Protein; 660 AA.
XX
AC   AAE07898;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #5.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium tetani.
OS   Chimeric - Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 46; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain fragment. This sequence is
CC   constructed by fusing the binding domain which is a hybrid of botulinum
CC   neurotoxin type F (BoNT/F) and tetanus neurotoxin (TeNT) domain II with
CC   translocation domain of diphtheria neurotoxin.
XX
SQ   Sequence   660 AA;


  Query Match           49.4%;  Score 213;  DB 22;  Length 660;
  Best Local Similarity  100.0%;  Pred. No. 5.7e-202;
  Matches  213;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 274

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    275 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 334

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 394

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYV 213
          |||||||||||||||||||||||||||||||||
Db    395 nlideksisnlgdihvsdnilfkivgcndtryv 427


RESULT  13
AAY77138
ID   AAY77138 standard; Protein; 432 AA.
XX
AC   AAY77138;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW   C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW   VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US15570.
XX
PR   10-JUL-1998;   98US-0092416.
PR   12-MAY-1999;   99US-0133870.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87216.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum -
XX
PS   Claim 27; Page -; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid
CC   fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC   serotypes A-G. In preferred embodiments of the invention, the vector is
CC   a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC   this vector results in the production of large amounts of a protein
CC   encoded by a sequence cloned into the replicon. The constructs are used
CC   to produce vaccines against botulism. The proteins can also be used as
CC   diagnostic tools for the diagnosis of botulism. The transformed host
CC   cells can be used to analyse the effectiveness of drugs and agents which
CC   inhibit toxin effects. The vaccine currently used against botulism is
CC   dangerous and expensive to produce, and contains formalin, which is very
CC   painful for the recipient. Also, the vaccine is incomplete, in that only
CC   5 of the 7 serotypes are represented in the formulation. The novel
CC   vaccine of overcomes these problems, as it is easily purified, and
CC   available in large quantities. It is also expressed in the lymph nodes
CC   for a better immune response. Sequences AAY77134-Y77139 represent
CC   synthetic BoNT Hc fragments used in the present invention. The DNA
```

```
CC   encoding these sequences had been optimised for codon usage for
CC   expression in yeast. Note: This sequence is not given in the
CC   specification, but is decoded from the BoNTF Hc DNA sequence given on
CC   pages 45-46.
XX
SQ   Sequence   432 AA;


  Query Match             49.2%;  Score 212;  DB 21;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-201;
  Matches  212;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      90 SFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLNYNKIIWTLQDTAGNNQKLVFNYT 149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      91 sfwvripkyfnkvnlnneytiidcirnnnsgwkislnynkiiwtlqdtagnnqklvfnyt 150

Qy     150 QMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKIVGCND 209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     151 qmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfkivgcnd 210

Qy     210 TRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLRTDKSI 269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     211 tryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllrtdksi 270

Qy     270 TQNSNFLNINQQRGVYQKPNIFSNTRLYTGVE 301
           ||||||||||||||||||||||||||||||||
Db     271 tqnsnflninqqrgvyqkpnifsntrlytgve 302


RESULT  14
AAW09015
ID   AAW09015 standard; Protein; 144 AA.
XX
AC   AAW09015;
XX
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin polypeptide (aa848-991).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB01409.
XX
PR   12-JUN-1995;   95GB-0011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows
PT   recombinant vaccine prodn.
XX
PS   Claim 5; Page 17-18; 37pp; English.
XX
CC   Novel polypeptides (AAW09014-17) respectively comprise amino acids
CC   848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a
CC   type F botulinum neurotoxin (BoNT/F).  They lack the L chain and
CC   HN epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   144 AA;


  Query Match             33.4%;  Score 144;  DB 18;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-134;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 60

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 120

Qy     121 WKISLNYNKIIWTLQDTAGNNQKL 144
           ||||||||||||||||||||||||
Db     121 wkislnynkiiwtlqdtagnnqkl 144


RESULT  15
AAW09016
ID   AAW09016 standard; Protein; 144 AA.
XX
AC   AAW09016;
XX
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin polypeptide (aa992-1135).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB01409.
XX
PR   12-JUN-1995;   95GB-0011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows
PT   recombinant vaccine prodn.
XX
PS   Claim 5; Page 18-19; 37pp; English.
XX
CC   Novel polypeptides (AAW09014-17) respectively comprise amino acids
CC   848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a
CC   type F botulinum neurotoxin (BoNT/F).  They lack the L chain and
CC   HN epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   144 AA;


  Query Match             33.4%;  Score 144;  DB 18;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-134;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     145 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 60
```

```
Qy     205 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 120

Qy     265 TDKSITQNSNFLNINQQRGVYQKP 288
           ||||||||||||||||||||||||
Db     121 tdksitqnsnflninqqrgvyqkp 144


Search completed: August 15, 2002, 11:12:25
Job time: 317 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 8, 2002, 09:39:06 ; Search time 96.3 Seconds
(without alignments)
497.122 Million cell updates/sec

Title: US-08-981-087A-1
Perfect score: 2288
Sequence: 1 SYTNDKILILYFNKLYKKIK.....TSSNGCFWSFISKEHGWQEN 431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2288	100.0	431	18	AAW09014	Immunogenic type F
2	2288	100.0	432	22	AAB04096	Botulism toxin hea
3	2288	100.0	432	22	AAB04103	Botulism toxin hea
4	2288	100.0	645	22	AAE07894	Modified clostridi
5	2288	100.0	685	22	AAE07893	Modified clostridi
6	2288	100.0	862	22	AAE07890	Modified clostridi
7	2288	100.0	887	22	AAE07892	Modified clostridi
8	2288	100.0	1032	22	AAE07901	C. botulinum C2 tr
9	2288	100.0	1059	21	AAY93309	A manganese supero
10	2288	100.0	1084	21	AAY93312	A manganese supero
11	2288	100.0	1092	22	AAE07900	C. botulinum C2 tr
12	2271	99.3	432	21	AAY77138	Synthetic botulinu
13	1800	78.7	448	19	AAW68399	Clostridium botuli
14	1451.5	63.4	449	21	AAY77137	Synthetic botulinu
15	1451.5	63.4	449	22	AAB04094	Botulism toxin hea
16	1447.5	63.3	452	19	AAW68396	Clostridium botuli
17	1426	62.3	451	19	AAW68395	Clostridium botuli
18	1423.5	62.2	419	22	AAB04095	Botulism toxin hea
19	1355.5	59.2	660	22	AAE07898	Modified clostridi
20	1120.5	49.0	1067	21	AAY93307	A manganese supero
21	1120.5	49.0	1092	21	AAY93310	A manganese supero
22	1120.5	49.0	1296	17	AAR95010	C. botulinum type
23	1113.5	48.7	437	22	AAB04088	Botulism toxin hea
24	1113.5	48.7	438	17	AAR95008	Type A neurotoxin
25	1113.5	48.7	438	19	AAW68389	Clostridium botuli
26	1113.5	48.7	438	21	AAY77134	Synthetic botulinu
27	1113.5	48.7	445	19	AAW68391	Clostridium botuli
28	1113.5	48.7	462	17	AAR95009	Type A neurotoxin
29	1113.5	48.7	462	19	AAW68390	Clostridium botuli
30	1111.5	48.6	434	22	AAB04089	Botulism toxin hea
31	1111.5	48.6	435	22	AAB04090	Botulism toxin hea
32	1106.5	48.4	837	21	AAY77140	Native botulinum n
33	1099.5	48.1	432	21	AAY77142	Native botulinum n
34	1088.5	47.6	847	22	AAB04081	Botulism toxin hea
35	1079.5	47.2	415	22	AAB04083	Botulism toxin C f
36	1019.5	44.6	382	21	AAB36303	BoNT/A prototoxin
37	772	33.7	472	19	AAW68393	Clostridium botuli
38	770	33.7	144	18	AAW09015	Immunogenic type F
39	769	33.6	1070	21	AAY93308	A manganese supero
40	769	33.6	1095	21	AAY93311	A manganese supero
41	769	33.6	1291	19	AAW68392	Clostridium botuli
42	761	33.3	144	18	AAW09016	Immunogenic type F
43	757	33.1	143	18	AAW09017	Immunogenic type F
44	751	32.8	472	19	AAW68394	Clostridium botuli
45	743	32.5	848	22	AAB04082	Botulism toxin hea

ALIGNMENTS

RESULT 1
AAW09014
ID AAW09014 standard; Protein; 431 AA.
XX
AC AAW09014;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX

```
CC   A polypeptide (AAW09014) comprises the heavy chain (amino acids
CC   848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be
CC   produced using a synthetic gene (AAT48101) based on the natural
CC   gene sequence (AAT48100) for the heavy chain.  The polypeptides and
CC   its fragments (see also AAW09015-17) lack the light chain and HN
CC   epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   431 AA;


  Query Match             100.0%;  Score 2288;  DB 18;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-168;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 60

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 120

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 180

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 240

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 300

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 360

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 420

Qy     421 FISKEHGWQEN 431
           |||||||||||
Db     421 fiskehgwqen 431


RESULT   2
AAB04096
ID   AAB04096 standard; Protein; 432 AA.
XX
AC   AAB04096;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54490.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Claim 3; Fig 9b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match             100.0%;  Score 2288;  DB 22;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 8e-168;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 181

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 241

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 301

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 361

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    362 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 421

Qy    421 FISKEHGWQEN 431
          ||||||||||||
Db    422 fiskehgwqen 432
```

RESULT 3
AAB04103
ID AAB04103 standard; Protein; 432 AA.
XX
AC AAB04103;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54499.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Disclosure; Fig 18b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 432 AA;

Query Match 100.0%; Score 2288; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 8e-168;

Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 181

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 241

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 301

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 361

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 421

Qy    421 FISKEHGWQEN 431
          ||||||||||||
Db    422 fiskehgwqen 432
```

RESULT 4
AAE07894
ID AAE07894 standard; Protein; 645 AA.
XX
AC AAE07894;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain fragment #1.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 44; 50pp; English.
XX

CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;

Query Match 100.0%; Score 2288; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 1.3e-167;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    215 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 274

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    275 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 334

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    335 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 394

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    395 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 454

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    455 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 514

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    515 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 574

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    575 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 634

Qy    421 FISKEHGWQEN 431
          ||||||||||||
Db    635 fiskehgwqen 645
```

RESULT 5
AAE07893
ID AAE07893 standard; Protein; 685 AA.
XX
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 43; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation peptide from influenza virus and a neuronal cell-specific
CC binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 685 AA;

Query Match 100.0%; Score 2288; DB 22; Length 685;
Best Local Similarity 100.0%; Pred. No. 1.4e-167;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 314

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 374

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    375 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 434

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 494

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 554

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    555 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 614
```

```
Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     615 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 674

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db     675 fiskehgwqen 685


RESULT   6
AAE07890
ID   AAE07890 standard; Protein; 862 AA.
XX
AC   AAE07890;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 40; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate.
CC   This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC   Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC   translocation domain from diphtheria neurotoxin and a neuronal cell-
CC   specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   862 AA;
```

```
  Query Match           100.0%;  Score 2288;  DB 22;  Length 862;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-167;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 491

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 551

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     552 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 611

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     612 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 671

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     672 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 731

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     732 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 791

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     792 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 851

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db     852 fiskehgwqen 862


RESULT   7
AAE07892
ID   AAE07892 standard; Protein; 887 AA.
XX
AC   AAE07892;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
```

```
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 42; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate.
CC   This conjugate comprises a mitochondrial leader sequence from human
CC   Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC   linker that can be cleaved by thrombin, translocation domain from
CC   diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC   botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   887 AA;

  Query Match             100.0%;  Score 2288;  DB 22;  Length 887;
  Best Local Similarity   100.0%;  Pred. No. 2e-167;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     457 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 516

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     517 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 576

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     577 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 636

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     637 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 696

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     697 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 756

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     757 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 816

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     817 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 876

Qy     421 FISKEHGWQEN 431
          ||||||||||||
Db     877 fiskehgwqen 887


RESULT   8
AAE07901
ID   AAE07901 standard; Protein; 1032 AA.
XX
AC   AAE07901;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 48; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1032 AA;

  Query Match             100.0%;  Score 2288;  DB 22;  Length 1032;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-167;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 661

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 721

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 781

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 841

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 901
```

```
Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 961

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     962 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1021

Qy     421 FISKEHGWQEN 431
           ||||||||||| 
Db    1022 fiskehgwqen 1032


RESULT   9
AAY93309
ID   AAY93309 standard; protein; 1059 AA.
XX
AC   AAY93309;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC   can be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a
CC   domain that translocates the SOD of the composition into the neuronal
CC   cell. After translocation, the linker is cleaved to release the SOD.
CC   The composition is useful for treating neuronal diseases caused or
CC   augmented by oxidative stress, such as ischemic stroke, trauma,
CC   Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ   Sequence   1059 AA;

  Query Match           100.0%;  Score 2288;  DB 21;  Length 1059;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-167;
  Matches  431;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     629 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 688

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     689 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 748

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     749 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 808

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     809 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 868

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     869 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 928

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     929 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 988

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     989 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1048

Qy     421 FISKEHGWQEN 431
           ||||||||||| 
Db    1049 fiskehgwqen 1059


RESULT  10
AAY93312
ID   AAY93312 standard; protein; 1084 AA.
XX
AC   AAY93312;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
```

PS Disclosure; Page 57-60; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide
CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC that can be cleaved by thrombin, and a heavy chain derived from
CC botulinum neurotoxin serotype F. The specification describes a
CC composition for delivery of SOD to neuronal cells. The composition
CC comprises SOD linked, by a cleavable linker, to a neuronal cell
CC targeting component (NCTC). This component has a domain that binds
CC to a neuronal cell and a domain that translocates the SOD of the
CC composition into the neuronal cell. After translocation, the linker
CC is cleaved to release the SOD. The composition is useful for treating
CC neuronal diseases caused or augmented by oxidative stress, such as
CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases.
XX
SQ Sequence 1084 AA;

Query Match 100.0%; Score 2288; DB 21; Length 1084;
Best Local Similarity 100.0%; Pred. No. 2.6e-167;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     654 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 713

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     714 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 773

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     774 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 833

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     834 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 893

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     894 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 953

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     954 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 1013

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1014 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1073

Qy     421 FISKEHGWQEN 431
          ||||||||||| 
Db    1074 fiskehgwqen 1084
```

RESULT 11
AAE07900
ID AAE07900 standard; Protein; 1092 AA.
XX
AC AAE07900;
XX
DT 01-NOV-2001 (first entry)
XX
DE C. botulinum C2 translocation domain with BoNT/F-binding domain #1.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Clostridium botulinum.
XX

PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 47; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC the invention.
XX
SQ Sequence 1092 AA;

ND

Query Match 100.0%; Score 2288; DB 22; Length 1092;
Best Local Similarity 100.0%; Pred. No. 2.6e-167;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 721

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 781

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     782 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 841

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     842 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 901

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     902 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 961

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     962 eviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpekiiklirtsnsn 1021

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1022 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 1081
```

```
Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db    1082 fiskehgwqen 1092


RESULT  12
AAY77138
ID   AAY77138 standard; Protein; 432 AA.
XX
AC   AAY77138;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW   C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW   VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US15570.
XX
PR   10-JUL-1998;   98US-0092416.
PR   12-MAY-1999;   99US-0133870.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87216.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum -
XX
PS   Claim 27; Page -; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid
CC   fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC   serotypes A-G. In preferred embodiments of the invention, the vector is
CC   a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC   this vector results in the production of large amounts of a protein
CC   encoded by a sequence cloned into the replicon. The constructs are used
CC   to produce vaccines against botulism. The proteins can also be used as
CC   diagnostic tools for the diagnosis of botulism. The transformed host
CC   cells can be used to analyse the effectiveness of drugs and agents which
CC   inhibit toxin effects. The vaccine currently used against botulism is
CC   dangerous and expensive to produce, and contains formalin, which is very
CC   painful for the recipient. Also, the vaccine is incomplete, in that only
CC   5 of the 7 serotypes are represented in the formulation. The novel
CC   vaccine of overcomes these problems, as it is easily purified, and
CC   available in large quantities. It is also expressed in the lymph nodes
CC   for a better immune response. Sequences AAY77134-Y77139 represent
CC   synthetic BoNT Hc fragments used in the present invention. The DNA
CC   encoding these sequences had been optimised for codon usage for
CC   expression in yeast. Note: This sequence is not given in the
CC   specification, but is decoded from the BoNTF Hc DNA sequence given on
CC   pages 45-46.
XX
SQ   Sequence   432 AA;

  Query Match             99.3%;  Score 2271;  DB 21;  Length 432;
  Best Local Similarity   99.3%;  Pred. No. 1.6e-166;
  Matches  428;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
Db      62 giysskpsevniaqnndiiyngryqnfsmsfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy     121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 wkislnynkiiwtlqdtagnnqklvfnytqmisisdyinkwifvtitnnrlgnsriying 181

Qy     181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 nlideksisnlgdihvsdnilfkivgcndtryvgiryfkvfdtelgkteietlysdepdp 241

Qy     241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 silkdfwgnyllynkryyllnllrtdksitqnsnflninqqrgvyqkpnifsntrlytgv 301

Qy     301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
           | ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     302 efiirkngstdisntdnfvrkndlayinvvdrdveyrlnadisiakpekiiklirtsnsn 361

Qy     361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 nslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynnirkntssngcfws 421

Qy     421 FISKEHGWQEN 431
           ||||||||||||
Db     422 fiskehgwqen 432


RESULT  13
AAW68399
ID   AAW68399 standard; Protein; 448 AA.
XX
AC   AAW68399;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type F toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide          1..21
FT                    /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US15394.
XX
PR   28-AUG-1996;   96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS,  Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
```

PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment
CC of Clostridium botulinum (202F strain) type F neurotoxin, encoded
CC by a DNA sequence (see AAV30593) in plasmid pETHisb. This vector
CC can be used to express BotC soluble C fragment in Escherichia
CC coli host cells, with the recombinant C fragment being purified on
CC an affinity column. The invention relates to recombinant proteins
CC derived from C. botulinum toxins, especially type B and type E
CC toxins. Methods are provided which allow for the isolation of
CC soluble recombinant proteins free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of
CC vaccines and antitoxins that are useful in the treatment of humans
CC and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 448 AA;

Query Match 78.7%; Score 1800; DB 19; Length 448;
Best Local Similarity 81.0%; Pred. No. 2.7e-130;
Matches 346; Conservative 33; Mismatches 40; Indels 8; Gaps 4;

```
Qy       7 ILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQFGIYSSK 66
           |||:|||:|||||||:|||||||||||||||||||||||||:|||||||||||||:|:
Db      27 iliiyfnrlykkikdssildmryennkfidisgygsnisingnvyiystnrnqfgiynsr 86

Qy      67 PSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSGWKISLN 126
           |||||||||||||| |||||||||||||::  :| | |||||:|: ||||||||||
Db      87 lsevniaqnndiiynsryqnfsisfwvripkhykpmnhnreytiincmgnnnsgwkislr 146

Qy     127 YNK---IIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLI 183
             :   ||||||||:|| : |:| | ::  ||:|||||||||||||||||||||||||||
Db     147 tvrdceiiwtlqdtsgnkenlifryeelnrisnyinkwifvtitnnrlgnsriyingnli 206

Qy     184 DEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSIL 243
           |||||||||||||||||||||:|  ||||||||||:||| ||||||||||:|||||||
Db     207 veksisnlgdihvsdnilfkivgcddetyvgiryfkvfnteldkteietlysnepdpsil 266

Qy     244 KDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGVEVI 303
           |::|||||||||:||| |||| || || ||  |||||||||  : ::| | :|| |||||
Db     267 knywgnyllynkkyylfnllrkdkyitlnsgilninqqrgv-tegsvflnyklyegvevi 325

Qy     304 IRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSNNSL 363
           |||||  |||||||||||||||||||||| |||||||    | ||  |:|||||  |:||
Db     326 irkngpidisntdnfvrkndlayinvvdrgveyrlyad---tksek-ekiirtsnlndsl 381

Qy     364 GQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWSFIS 423
           ||||||||||||||||||||| ||||||||||||||||||||:|||||||||| ||
Db     382 gqiivmdsignnctmnfqnnngsnigllgfhsnnlvasswyynnirrntssngcfwssis 441

Qy     424 KEHGWQE 430
           ||:||:|
Db     442 kengwke 448
```

RESULT 14
AAY77137
ID AAY77137 standard; Protein; 449 AA.
XX
AC AAY77137;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype E (BoNTE) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype E;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87215.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
PS Claim 26; Page 43-44; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAY77134-Y77139 represent
CC synthetic BoNT Hc fragments used in the present invention. The DNA
CC encoding these sequences had been optimised for codon usage for
CC expression in yeast.
XX
SQ Sequence 449 AA;

Query Match 63.4%; Score 1451.5; DB 21; Length 449;
Best Local Similarity 63.0%; Pred. No. 1.7e-103;
Matches 274; Conservative 74; Mismatches 70; Indels 17; Gaps 7;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           |||:||||| |||| :|:|| :|:|:|||:|:|::| ||| |||:|||||| | ||:|||
Db      26 sytddkilisyfnkffkriksssvlnmrykndkyvdtsgydsniningdvykyptnknqf 85

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNS 119
           |||: | :|:||:||: |||: :|:|||||||||| | || ||:|||||||:|:|:|||
Db      86 giyndkltelnisqndyiiydnkyknfsisfwvripnydnkivnvnneytiincmrdnns 145

Qy     120 GWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYIN 179
           |||:|||:|:|||||| || |||| |||     |||||||||||||||:|||:|::|||
Db     146 gwkvslnhneiiwtlqdnaginqklafnygnangisdyinkwifvtitndrlgdsklyin 205

Qy     180 GNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPD 239
           |||||:||| |||:||||||||||||| |: |||:||||| :|| || :|||:||||:||:
Db     206 gnlidqksilnlgnihvsdnilfkivncsytryigiryfnifdkeldeteiqtlysnepn 265

Qy     240 PSILKDFWGNYLLYNKRYYLLNLLRTDKSI-TQNSNFLNINQQRGVYQKPNIFSNTRLYT 298
           :|||||||||||:| |||||:|: :  |  :  : |:||  |       |    |||:
```

```
Db    266 tnilkdfwgnyllydkeyyllnvlkpnnfidrrkdstlsinnirs-----tillanrlys 320

Qy    299 GVEVIIRK--NGSTDISNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIR 355
          |::| |::  | ||:    || |||||  ||| |  :   : |||| :    || ||:
Db    321 gikvkiqrvnnsstn----dnlvrkndqvyinfvaskthlfplyadtattnkektiki-- 374

Qy    356 TSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSN 415
           |:| |   |::||:|:|||||||:|||| ||||||| :: :|||:||| ::| :|:||
Db    375 -sssgnrfnqvvvmnsvgnnctmnfknnngnnigllgfkadtvvastwyythmrdhtnsn 433

Qy    416 GCFWSFISKEHGWQE 430
          ||||:|||:||||||
Db    434 gcfwnfiseehgwqe 448


RESULT 15
AAB04094
ID   AAB04094 standard; Protein; 449 AA.
XX
AC   AAB04094;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54488.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Claim 3; Fig 7b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   449 AA;


  Query Match             63.4%;  Score 1451.5;  DB 22;  Length 449;
  Best Local Similarity   63.0%;  Pred. No. 1.7e-103;
  Matches  274;  Conservative   74;  Mismatches   70;  Indels   17;  Gaps    7;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          |||:||||| |||| :|:|| :|:|:|||:|:|::| ||| |||:|||||| | ||:|||
Db     26 sytddkilisyfnkffkriksssvlnmrykndkyvdtsgydsniningdvykyptnknqf 85

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNK-VNLNNEYTIIDCIRNNNS 119
          |||: | :|:||:||: |||: :|:|||||||||||| | || ||:|||||||:|:|:|||
Db     86 giyndkltelnisqndyiiydnkyknfsisfwvripnydnkivnvnneytiincmrdnns 145

Qy    120 GWKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYIN 179
          |||:|||:|:|||||| || |||| |||     |||||||||||||||:|||:|::|||
Db    146 gwkvslnhneiiwtlqdnaginqklafnygnangisdyinkwifvtitndrlgdsklyin 205

Qy    180 GNLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPD 239
          |||||:||| |||:||||||||||||| |: |||:||||| :|| || :|||:||||:||:
Db    206 gnlidqksilnlgnihvsdnilfkivncsytryigiryfnifdkeldeteiqtlysnepn 265

Qy    240 PSILKDFWGNYLLYNKRYYLLNLLRTDKSI-TQNSNFLNINQQRGVYQKPNIFSNTRLYT 298
          :|||||||||||:| |||||:|: :  |  :  : |:||  |      |   |||:
Db    266 tnilkdfwgnyllydkeyyllnvlkpnnfidrrkdstlsinnirs-----tillanrlys 320

Qy    299 GVEVIIRK--NGSTDISNTDNFVRKNDLAYIN-VVDRDVEYRLYADISIAKPEKIIKLIR 355
          |::| |::  | ||:    || |||||  ||| |  :   : |||| :    || ||:
Db    321 gikvkiqrvnnsstn----dnlvrkndqvyinfvaskthlfplyadtattnkektiki-- 374

Qy    356 TSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSN 415
           |:| |   |::||:|:|||||||:|||| ||||||| :: :|||:||| ::| :|:||
Db    375 -sssgnrfnqvvvmnsvgnnctmnfknnngnnigllgfkadtvvastwyythmrdhtnsn 433

Qy    416 GCFWSFISKEHGWQE 430
          ||||:|||:||||||
Db    434 gcfwnfiseehgwqe 448


Search completed: August  8, 2002, 09:42:45
Job time: 219 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:13:13 ; Search time 24.69 Seconds
(without alignments)
675.906 Million cell updates/sec

Title: US-08-981-087A-1
Perfect score: 431
Sequence: 1 SYTNDKILILYFNKLYKKIK.....TSSNGCFWSFISKEHGWQEN 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	6.3	1274	1	BXF_CLOBO	P30996 clostridium
2	15	3.5	1250	1	BXE_CLOBO	Q00496 clostridium
3	15	3.5	1250	1	BXE_CLOBU	P30995 clostridium
4	11	2.6	1295	1	BXA1_CLOBO	P10845 clostridium
5	11	2.6	1295	1	BXA2_CLOBO	Q45894 clostridium
6	10	2.3	458	1	MB22_ARATH	O80950 arabidopsis
7	8	1.9	1290	1	BXB_CLOBO	P10844 clostridium
8	8	1.9	1296	1	BXG_CLOBO	Q60393 clostridium
9	7	1.6	199	1	TDX2_THEAC	Q9hjl3 thermoplasm
10	7	1.6	267	1	EUTT_ECOLI	P76554 escherichia
11	7	1.6	267	1	EUTT_SALTY	Q9zfv4 salmonella
12	7	1.6	358	1	VAL1_CLVK	P14982 cassava lat
13	7	1.6	358	1	VAL1_CLVN	P14972 cassava lat
14	7	1.6	387	1	FIB2_ADE40	P18048 human adeno
15	7	1.6	387	1	FIB2_ADE41	P16883 human adeno
16	7	1.6	432	1	YZ21_METJA	Q60281 methanococc
17	7	1.6	441	1	MGTA_THEMA	P80099 thermotoga
18	7	1.6	449	1	MURF_RICPR	O05953 rickettsia
19	7	1.6	501	1	VGLC_HSVMB	P22650 marek's dis
20	7	1.6	501	1	VGLC_HSVMD	P33500 marek's dis
21	7	1.6	501	1	VGLC_HSVMM	P22651 marek's dis
22	7	1.6	505	1	VGLC_HSVMG	P10681 marek's dis
23	7	1.6	508	1	YM05_ARCFU	O28078 archaeoglob
24	7	1.6	511	1	RRB1_YEAST	Q04225 saccharomyc
25	7	1.6	515	1	CRBA_DROME	P29747 drosophila
26	7	1.6	523	1	NAM8_YEAST	Q00539 saccharomyc
27	7	1.6	608	1	EDD_HELPJ	Q9zkb3 helicobacte
28	7	1.6	608	1	EDD_HELPY	P56111 helicobacte
29	7	1.6	665	1	ENV_MLVMO	P03385 moloney mur
30	7	1.6	781	1	PARC_MYCGE	P47446 mycoplasma
31	7	1.6	814	1	PI3K_ARATH	P42339 arabidopsis
32	7	1.6	942	1	AMPN_MANSE	P91885 manduca sex
33	7	1.6	944	1	Y166_UREPA	Q9pqx7 ureaplasma
34	7	1.6	1035	1	ITA9_HUMAN	Q13797 homo sapien
35	7	1.6	1132	1	PHY1_PHYPA	P36505 physcomitre
36	7	1.6	1139	1	HMW1_MYCGE	Q49413 mycoplasma
37	7	1.6	1276	1	BXD_CLOBO	P19321 clostridium
38	7	1.6	1655	1	OMPB_RICCN	Q9kka3 r outer mem
39	7	1.6	1656	1	OMPB_RICJA	O06653 r outer mem
40	7	1.6	2292	1	POLG_EMCVB	P17593 encephalomy
41	7	1.6	2292	1	POLG_EMCVD	P17594 encephalomy
42	6	1.4	48	1	ATP8_ASPAM	P00858 aspergillus
43	6	1.4	48	1	ATP8_EMENI	P00857 emericella
44	6	1.4	49	1	Y195_BPT7	P03804 bacteriopha
45	6	1.4	63	1	RL29_BUCAK	P46174 buchnera ap

ALIGNMENTS

RESULT 1
BXF_CLOBO
ID BXF_CLOBO STANDARD; PRT; 1274 AA.
AC P30996;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE (Bontoxilysin F).
GN BOTF.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23387;
RX MEDLINE=93012902; PubMed=1398040;
RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA Roberts T.A., Thompson D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum.";
RL FEMS Microbiol. Lett. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=HOBBS FT10;
RX MEDLINE=94297488; PubMed=7764998;
RA East A.K., Collins M.D.;
RT "Conserved structure of genes encoding components of botulinum
RT neurotoxin complex M and the sequence of the gene coding for the
RT nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL Curr. Microbiol. 29:69-77(1994).
RN [3]
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC BOND OF SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No

```
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;


  Query Match             6.3%;  Score 27;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-19;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    157 YINKWIFVTITNNRLGNSRIYINGNLI 183
          ||||||||||||||||||||||||||
Db   1006 YINKWIFVTITNNRLGNSRIYINGNLI 1032


RESULT   2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BELUGA;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; A60027; A60027.
```

```
DR   PIR; B35294; B35294.
DR   PIR; JH0257; JH0257.
DR   PIR; S08575; S08575.
DR   PIR; S18111; S18111.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;


  Query Match             3.5%;  Score 15;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 ISDYINKWIFVTITN 168
           |||||||||||||||
Db     982 ISDYINKWIFVTITN 996


RESULT   3
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, AND ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   PIR; S16145; S16145.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;


  Query Match             3.5%;  Score 15;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 ISDYINKWIFVTITN 168
           |||||||||||||||
Db     982 ISDYINKWIFVTITN 996


RESULT   4
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P18639; P01561;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
```

```
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=HALL;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-!-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX(R) (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PIR; S09492; S09492.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;

  Query Match             2.6%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.0065;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     161 WIFVTITNNRL 171
           ||||||||||| 
Db    1013 WIFVTITNNRL 1023


RESULT   5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;

  Query Match             2.6%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.0065;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     161 WIFVTITNNRL 171
           ||||||||||| 
Db    1013 WIFVTITNNRL 1023
```

```
RESULT   6
MB22_ARATH
ID   MB22_ARATH     STANDARD;      PRT;   458 AA.
AC   O80950;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Myrosinase binding protein-like At2g39310.
GN   AT2G39310 OR T16B24.5.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. COLUMBIA;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
CC   -!- SIMILARITY: BELONGS TO THE JACALIN LECTIN FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AC004697; AAC28979.1; -.
DR   HSSP; P18674; 1JOT.
DR   InterPro; IPR001229; Jacalin.
DR   Pfam; PF01419; Jacalin; 3.
KW   Lectin; Repeat; Multigene family.
SQ   SEQUENCE   458 AA;  50463 MW;  EB01A410563EAAF8 CRC64;


  Query Match             2.3%;  Score 10;  DB 1;  Length 458;
  Best Local Similarity   100.0%;  Pred. No. 0.027;
  Matches   10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     285 YQKPNIFSNT 294
           ||||||||||
Db     238 YQKPNIFSNT 247


RESULT   7
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; S07128; S07128.
DR   PIR; S07155; S07155.
DR   PIR; S08562; S08562.
DR   PIR; S08573; S08573.
DR   PIR; S08574; S08574.
DR   PIR; A48940; A48940.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;


  Query Match             1.9%;  Score 8;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 7.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     957 NNSGWKIS 964


RESULT   8
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;


  Query Match             1.9%;  Score 8;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 ISDYINKW 161
           ||||||||
Db    1001 ISDYINKW 1008


RESULT   9
TDX2_THEAC
ID   TDX2_THEAC     STANDARD;      PRT;   199 AA.
AC   Q9HJL3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable peroxiredoxin 2.
GN   TA0954.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
CC   -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
CC   ----------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL445066; CAC12083.1; -.
DR   InterPro; IPR000866; AhpC-TSA.
DR   Pfam; PF00578; AhpC-TSA; 1.
KW   Antioxidant; Complete proteome.
FT   ACT_SITE     40     40       BY SIMILARITY.
SQ   SEQUENCE   199 AA;  22581 MW;  4E266162F5D58162 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 199;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     181 NLIDEKS 187
           |||||||
Db     104 NLIDEKS 110


RESULT  10
EUTT_ECOLI
ID   EUTT_ECOLI     STANDARD;      PRT;   267 AA.
AC   P76554;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).
GN   EUTT OR B2459.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + cob(I)alamin + H(2)O = phosphate +
CC       diphosphate + adenosylcobalamin.
CC   -!- PATHWAY: ETHANOLAMINE UTILIZATION.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE000332; AAC75512.1; -.
DR   EcoGene; EG14189; eutT.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   267 AA;  30171 MW;  E51EDAB528B4FA76 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     259 LLNLLRT 265
           |||||||
Db     199 LLNLLRT 205


RESULT  11
EUTT_SALTY
ID   EUTT_SALTY     STANDARD;      PRT;   267 AA.
AC   Q9ZFV4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).
GN   EUTT OR STM2467.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=99395039; PubMed=10464203;
RA   Kofoid E.C., Rappleye C.A., Stojiljkovic I., Roth J.R.;
RT   "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium
RT   encodes five homologues of carboxysome shell proteins.";
RL   J. Bacteriol. 181:5317-5329(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + cob(I)alamin + H(2)O = phosphate +
CC       diphosphate + adenosylcobalamin.
CC   -!- PATHWAY: ETHANOLAMINE UTILIZATION.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF093749; AAC78114.1; -.
DR   EMBL; AE008811; AAL21361.1; -.
DR   StyGene; SG10636; eutT.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   267 AA;  30238 MW;  9502A28FDB4DC9E4 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     259 LLNLLRT 265
           |||||||
Db     199 LLNLLRT 205


RESULT  12
VAL1_CLVK
ID   VAL1_CLVK      STANDARD;      PRT;   358 AA.
AC   P14982;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   AL1 protein (40.4 kDa protein).
```

```
GN   AC1.
OS   Cassava latent virus (strain West Kenyan 844).
OC   Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX   NCBI_TaxID=10818;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stanley J., Gay M.R.;
RT   "Nucleotide sequence of cassava latent virus DNA.";
RL   Nature 301:260-262(1983).
CC   -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02057; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR001191; Gemini_AL1.
DR   Pfam; PF00799; Gemini_AL1; 1.
DR   PRINTS; PR00227; GEMCOATAL1.
DR   ProDom; PD000736; Gemini_AL1; 1.
KW   ATP-binding.
FT   NP_BIND     220    227       ATP (POTENTIAL).
SQ   SEQUENCE   358 AA;  40346 MW;  ED173E753EE92D69 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     165 TITNNRL 171
           |||||||
Db      68 TITNNRL 74


RESULT  13
VAL1_CLVN
ID   VAL1_CLVN      STANDARD;      PRT;   358 AA.
AC   P14972;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   AL1 protein (40.4 kDa protein).
GN   AC1.
OS   Cassava latent virus (strain Nigerian).
OC   Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX   NCBI_TaxID=10819;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90174930; PubMed=2308831;
RA   Morris B., Coates L., Lowe S., Richardson K., Eddy P.;
RT   "Nucleotide sequence of the infectious cloned DNA components of
RT   African cassava mosaic virus (Nigerian strain).";
RL   Nucleic Acids Res. 18:197-198(1990).
CC   -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17095; CAA34953.1; -.
DR   PIR; S07594; S07594.
DR   InterPro; IPR001191; Gemini_AL1.
DR   Pfam; PF00799; Gemini_AL1; 1.
DR   PRINTS; PR00227; GEMCOATAL1.
DR   ProDom; PD000736; Gemini_AL1; 1.
KW   ATP-binding.
FT   NP_BIND     220    227       ATP (POTENTIAL).
SQ   SEQUENCE   358 AA;  40435 MW;  1DB16BB0CB2D5E2C CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     165 TITNNRL 171
           |||||||
Db      68 TITNNRL 74


RESULT  14
FIB2_ADE40
ID   FIB2_ADE40     STANDARD;      PRT;   387 AA.
AC   P18048;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Fiber protein 2.
OS   Human adenovirus type 40.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=28284;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DUGAN;
RX   MEDLINE=94087748; PubMed=8263936;
RA   Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT   "The DNA sequence of adenovirus type 40.";
RL   J. Mol. Biol. 234:1308-1316(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93297140; PubMed=8517033;
RA   Kidd A.H., Chroboczek J., Cusack S., Ruigrok R.W.H.;
RT   "Adenovirus type 40 virions contain two distinct fibers.";
RL   Virology 192:73-84(1993).
RN   [3]
RP   SEQUENCE OF 167-387 FROM N.A.
RX   MEDLINE=89370295; PubMed=2773314;
RA   Kidd A.H., Erasmus M.J.;
RT   "Sequence characterization of the adenovirus 40 fiber gene.";
RL   Virology 172:134-144(1989).
CC   -!- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC       BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L19443; AAC13979.1; -.
DR   EMBL; M28822; AAA03233.1; -.
DR   PIR; A40048; ERADY4.
DR   InterPro; IPR000939; Adeno_fiber2.
DR   InterPro; IPR000978; Adeno_fiber_knob.
DR   InterPro; IPR000931; Adeno_fibre.
DR   Pfam; PF00541; adeno_fiber; 1.
DR   Pfam; PF00608; adeno_fiber2; 5.
DR   PRINTS; PR00307; ADENOVSFIBRE.
KW   Fiber protein.
FT   CONFLICT    226    226       G -> S (IN REF. 2 AND 3).
SQ   SEQUENCE   387 AA;  41346 MW;  11A3C1FCD61A3ACB CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy     358 NSNNSLG 364
           |||||||
Db      89 NSNNSLG 95


RESULT  15
FIB2_ADE41
ID   FIB2_ADE41     STANDARD;      PRT;   387 AA.
AC   P16883;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Fiber protein 2.
OS   Human adenovirus type 41.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=10524;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TAK;
RX   MEDLINE=90245595; PubMed=2336370;
RA   Pieniazek N.J., Slemenda S.B., Pieniazek D., Velarde J. Jr.,
RA   Luftig R.B.;
RT   "Human enteric adenovirus type 41 (Tak) contains a second fiber
RT   protein gene.";
RL   Nucleic Acids Res. 18:1901-1901(1990).
RN   [2]
RP   SEQUENCE OF 337-387 FROM N.A.
RC   STRAIN=FB585;
RX   MEDLINE=91021015; PubMed=2219717;
RA   Kidd A.H., Erasmus M.J., Tiemessen C.T.;
RT   "Fiber sequence heterogeneity in subgroup F adenoviruses.";
RL   Virology 179:139-150(1990).
CC   -!- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC       BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X17016; CAA34882.1; -.
DR   EMBL; M60327; AAA42505.1; -.
DR   PIR; S09217; ERADN1.
DR   PIR; A45352; A45352.
DR   HSSP; P11818; 1KNB.
DR   InterPro; IPR000939; Adeno_fiber2.
DR   InterPro; IPR000978; Adeno_fiber_knob.
DR   InterPro; IPR000931; Adeno_fibre.
DR   Pfam; PF00541; adeno_fiber; 1.
DR   Pfam; PF00608; adeno_fiber2; 5.
DR   PRINTS; PR00307; ADENOVSFIBRE.
KW   Fiber protein.
SQ   SEQUENCE   387 AA;  41397 MW;  8652E785276573C7 CRC64;


  Query Match             1.6%;  Score 7;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     358 NSNNSLG 364
           |||||||
Db      89 NSNNSLG 95


Search completed: August 15, 2002, 11:24:37
Job time: 684 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:25 ; Search time 96.53 Seconds
(without alignments)
165.696 Million cell updates/sec

Title: US-08-981-087A-2
Perfect score: 144
Sequence: 1 SYTNDKILILYFNKLYKKIK.....LNYNKIIWTLQDTAGNNQKL 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144	100.0	144	18	AAW09015	Immunogenic type F
2	144	100.0	431	18	AAW09014	Immunogenic type F
3	144	100.0	432	22	AAB04096	Botulism toxin hea
4	144	100.0	432	22	AAB04103	Botulism toxin hea
5	144	100.0	645	22	AAE07894	Modified clostridi
6	144	100.0	660	22	AAE07898	Modified clostridi
7	144	100.0	685	22	AAE07893	Modified clostridi
8	144	100.0	862	22	AAE07890	Modified clostridi
9	144	100.0	887	22	AAE07892	Modified clostridi
10	144	100.0	1032	22	AAE07901	C. botulinum C2 tr
11	144	100.0	1059	21	AAY93309	A manganese supero
12	144	100.0	1084	21	AAY93312	A manganese supero
13	144	100.0	1092	22	AAE07900	C. botulinum C2 tr
14	88	61.1	432	21	AAY77138	Synthetic botulinu
15	26	18.1	448	19	AAW68399	Clostridium botuli
16	11	7.6	419	22	AAB04095	Botulism toxin hea
17	11	7.6	449	21	AAY77137	Synthetic botulinu
18	11	7.6	449	22	AAB04094	Botulism toxin hea
19	11	7.6	451	19	AAW68395	Clostridium botuli
20	11	7.6	452	19	AAW68396	Clostridium botuli
21	8	5.6	233	21	AAY77143	Botulinum neurotox
22	8	5.6	382	21	AAB36303	BoNT/A prototoxin
23	8	5.6	415	22	AAB04083	Botulism toxin C f
24	8	5.6	432	21	AAY77142	Native botulinum n
25	8	5.6	434	22	AAB04089	Botulism toxin hea
26	8	5.6	435	22	AAB04090	Botulism toxin hea
27	8	5.6	437	22	AAB04088	Botulism toxin hea
28	8	5.6	438	17	AAR95008	Type A neurotoxin
29	8	5.6	438	19	AAW68389	Clostridium botuli
30	8	5.6	438	21	AAY77134	Synthetic botulinu
31	8	5.6	439	22	AAB04085	Botulism toxin C f
32	8	5.6	440	21	AAY77135	Synthetic botulinu
33	8	5.6	440	22	AAB04091	Botulism toxin hea
34	8	5.6	445	19	AAW68391	Clostridium botuli
35	8	5.6	462	17	AAR95009	Type A neurotoxin
36	8	5.6	462	19	AAW68390	Clostridium botuli
37	8	5.6	472	19	AAW68394	Clostridium botuli
38	8	5.6	472	19	AAW68393	Clostridium botuli
39	8	5.6	837	21	AAY77140	Native botulinum n
40	8	5.6	847	22	AAB04081	Botulism toxin hea
41	8	5.6	848	22	AAB04082	Botulism toxin hea
42	8	5.6	1067	21	AAY93307	A manganese supero
43	8	5.6	1070	21	AAY93308	A manganese supero
44	8	5.6	1092	21	AAY93310	A manganese supero
45	8	5.6	1095	21	AAY93311	A manganese supero

ALIGNMENTS

RESULT 1
AAW09015
ID AAW09015 standard; Protein; 144 AA.
XX
AC AAW09015;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin polypeptide (aa848-991).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 17-18; 37pp; English.
XX
CC Novel polypeptides (AAW09014-17) respectively comprise amino acids

```
CC   848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a
CC   type F botulinum neurotoxin (BoNT/F).  They lack the L chain and
CC   HN epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   144 AA;


  Query Match             100.0%;  Score 144;  DB 18;  Length 144;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-141;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 60

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 120

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    121 wkislnynkiiwtlqdtagnnqkl 144


RESULT   2
AAW09014
ID   AAW09014 standard; Protein; 431 AA.
XX
AC   AAW09014;
XX
DT   31-MAR-1997  (first entry)
XX
DE   Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW   Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS   Clostridium botulinum type F strain Langeland.
XX
PN   WO9641881-A1.
XX
PD   27-DEC-1996.
XX
PF   12-JUN-1996;   96WO-GB01409.
XX
PR   12-JUN-1995;   95GB-0011909.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Elmore MJ,  Mauchline ML,  Minton NP,  Pasechnik VA;
XX
DR   WPI; 1997-065467/06.
DR   N-PSDB; AAT48100.
XX
PT   Immunogenic type F botulinum toxin polypeptide(s) - allows
PT   recombinant vaccine prodn.
XX
PS   Claim 5; Page 16-17; 37pp; English.
XX
CC   A polypeptide (AAW09014) comprises the heavy chain (amino acids
CC   848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be
CC   produced using a synthetic gene (AAT48101) based on the natural
CC   gene sequence (AAT48100) for the heavy chain.  The polypeptides and
CC   its fragments (see also AAW09015-17) lack the light chain and HN
CC   epitopes necessary for metalloprotease activity and toxin
CC   internalisation.  They are free of botulinum toxin activity but can
CC   induce protective immunity to a type F botulinum toxin, making them
CC   useful for vaccine prodn.  Recombinant polypeptides can be
CC   produced in transformed host cells, esp. as fusion proteins, e.g.
CC   with maltose binding protein to facilitate purification.
XX
SQ   Sequence   431 AA;


  Query Match             100.0%;  Score 144;  DB 18;  Length 431;
  Best Local Similarity   100.0%;  Pred. No. 7.6e-141;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 60

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 120

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    121 wkislnynkiiwtlqdtagnnqkl 144


RESULT   3
AAB04096
ID   AAB04096 standard; Protein; 432 AA.
XX
AC   AAB04096;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54490.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Claim 3; Fig 9b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
```

```
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match           100.0%;  Score 144;  DB 22;  Length 432;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-141;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    122 wkislnynkiiwtlqdtagnnqkl 145


RESULT   4
AAB04103
ID   AAB04103 standard; Protein; 432 AA.
XX
AC   AAB04103;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54499.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Disclosure; Fig 18b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match           100.0%;  Score 144;  DB 22;  Length 432;
  Best Local Similarity  100.0%;  Pred. No. 7.6e-141;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 121

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    122 wkislnynkiiwtlqdtagnnqkl 145


RESULT   5
AAE07894
ID   AAE07894 standard; Protein; 645 AA.
XX
AC   AAE07894;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
```

XX
PS Example 2; Page 44; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;

Query Match 100.0%; Score 144; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 1.1e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
Db 215 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 274

Qy 61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
Db 275 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 334

Qy 121 WKISLNYNKIIWTLQDTAGNNQKL 144
Db 335 wkislnynkiiwtlqdtagnnqkl 358

RESULT 6
AAE07898
ID AAE07898 standard; Protein; 660 AA.
XX
AC AAE07898;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain fragment #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium tetani.
OS Chimeric - Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 46; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain which is a hybrid of botulinum
CC neurotoxin type F (BoNT/F) and tetanus neurotoxin (TeNT) domain II with
CC translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 660 AA;

Query Match 100.0%; Score 144; DB 22; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.1e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
Db 215 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 274

Qy 61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
Db 275 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 334

Qy 121 WKISLNYNKIIWTLQDTAGNNQKL 144
Db 335 wkislnynkiiwtlqdtagnnqkl 358

RESULT 7
AAE07893
ID AAE07893 standard; Protein; 685 AA.
XX
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX

PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 43; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation peptide from influenza virus and a neuronal cell-specific
CC binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 685 AA;

Query Match 100.0%; Score 144; DB 22; Length 685;
Best Local Similarity 100.0%; Pred. No. 1.2e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     255 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 314

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     315 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 374

Qy     121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db     375 wkislnynkiiwtlqdtagnnqkl 398
```

RESULT 8
AAE07890
ID AAE07890 standard; Protein; 862 AA.
XX
AC AAE07890;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 40; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate.
CC This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation domain from diphtheria neurotoxin and a neuronal cell-
CC specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 862 AA;

Query Match 100.0%; Score 144; DB 22; Length 862;
Best Local Similarity 100.0%; Pred. No. 1.4e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 491

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 551

Qy     121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db     552 wkislnynkiiwtlqdtagnnqkl 575
```

RESULT 9
AAE07892
ID AAE07892 standard; Protein; 887 AA.
XX
AC AAE07892;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW botulinum neurotoxin type F; BoNT/F.
XX

OS Chimeric - Homo sapiens.
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 42; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate.
CC This conjugate comprises a mitochondrial leader sequence from human
CC Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC linker that can be cleaved by thrombin, translocation domain from
CC diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 887 AA;

Query Match 100.0%; Score 144; DB 22; Length 887;
Best Local Similarity 100.0%; Pred. No. 1.5e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 516

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 576

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    577 wkislnynkiiwtlqdtagnnqkl 600
```

RESULT 10
AAE07901
ID AAE07901 standard; Protein; 1032 AA.
XX
AC AAE07901;
XX
DT 01-NOV-2001 (first entry)
XX
DE C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 48; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC the invention.
XX
SQ Sequence 1032 AA;

Query Match 100.0%; Score 144; DB 22; Length 1032;
Best Local Similarity 100.0%; Pred. No. 1.7e-140;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 661

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 721

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    722 wkislnynkiiwtlqdtagnnqkl 745
```

RESULT 11
AAY93309
ID AAY93309 standard; protein; 1059 AA.
XX
AC AAY93309;

```
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC   can be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a
CC   domain that translocates the SOD of the composition into the neuronal
CC   cell. After translocation, the linker is cleaved to release the SOD.
CC   The composition is useful for treating neuronal diseases caused or
CC   augmented by oxidative stress, such as ischemic stroke, trauma,
CC   Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ   Sequence   1059 AA;


  Query Match           100.0%;  Score 144;  DB 21;  Length 1059;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-140;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    629 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 688

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    689 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 748

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    749 wkislnynkiiwtlqdtagnnqkl 772


RESULT  12
AAY93312
ID   AAY93312 standard; protein; 1084 AA.
XX
AC   AAY93312;
```

```
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 57-60; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide
CC   dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC   that can be cleaved by thrombin, and a heavy chain derived from
CC   botulinum neurotoxin serotype F. The specification describes a
CC   composition for delivery of SOD to neuronal cells. The composition
CC   comprises SOD linked, by a cleavable linker, to a neuronal cell
CC   targeting component (NCTC). This component has a domain that binds
CC   to a neuronal cell and a domain that translocates the SOD of the
CC   composition into the neuronal cell. After translocation, the linker
CC   is cleaved to release the SOD. The composition is useful for treating
CC   neuronal diseases caused or augmented by oxidative stress, such as
CC   ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases.
XX
SQ   Sequence   1084 AA;


  Query Match           100.0%;  Score 144;  DB 21;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-140;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    654 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 713

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    714 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 773

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    774 wkislnynkiiwtlqdtagnnqkl 797


RESULT  13
AAE07900
```

```
ID   AAE07900 standard; Protein; 1092 AA.
XX
AC   AAE07900;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 47; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1092 AA;


  Query Match           100.0%;  Score 144;  DB 22;  Length 1092;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-140;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 721

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 giysskpsevniaqnndiiyngryqnfsisfwvripkyfnkvnlnneytiidcirnnnsg 781

Qy    121 WKISLNYNKIIWTLQDTAGNNQKL 144
          |||||||||||||||||||||||
Db    782 wkislnynkiiwtlqdtagnnqkl 805


RESULT  14
```

```
AAY77138
ID   AAY77138 standard; Protein; 432 AA.
XX
AC   AAY77138;
XX
DT   08-MAY-2000  (first entry)
XX
DE   Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW   Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW   C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW   VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS   Clostridium botulinum.
OS   Synthetic.
XX
PN   WO200002524-A2.
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US15570.
XX
PR   10-JUL-1998;   98US-0092416.
PR   12-MAY-1999;   99US-0133870.
XX
PA   (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI   Lee JS,  Pushko P,  Smith JF,  Parker M,  Dertzbaugh MT,  Smith L;
XX
DR   WPI; 2000-160827/14.
DR   N-PSDB; AAZ87216.
XX
PT   Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT   toxin serotypes A-G, is used for inducing an immune response against
PT   botulinum -
XX
PS   Claim 27; Page -; 54pp; English.
XX
CC   The invention relates to novel vaccines that induce a protective immune
CC   response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC   and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC   DNA construct comprising a vector, and at least one nucleic acid
CC   fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC   serotypes A-G. In preferred embodiments of the invention, the vector is
CC   a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC   this vector results in the production of large amounts of a protein
CC   encoded by a sequence cloned into the replicon. The constructs are used
CC   to produce vaccines against botulism. The proteins can also be used as
CC   diagnostic tools for the diagnosis of botulism. The transformed host
CC   cells can be used to analyse the effectiveness of drugs and agents which
CC   inhibit toxin effects. The vaccine currently used against botulism is
CC   dangerous and expensive to produce, and contains formalin, which is very
CC   painful for the recipient. Also, the vaccine is incomplete, in that only
CC   5 of the 7 serotypes are represented in the formulation. The novel
CC   vaccine of overcomes these problems, as it is easily purified, and
CC   available in large quantities. It is also expressed in the lymph nodes
CC   for a better immune response. Sequences AAY77134-Y77139 represent
CC   synthetic BoNT Hc fragments used in the present invention. The DNA
CC   encoding these sequences had been optimised for codon usage for
CC   expression in yeast. Note: This sequence is not given in the
CC   specification, but is decoded from the BoNTF Hc DNA sequence given on
CC   pages 45-46.
XX
SQ   Sequence   432 AA;


  Query Match           61.1%;  Score 88;  DB 21;  Length 432;
  Best Local Similarity  100.0%;  Pred. No. 8.1e-83;
  Matches   88;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 sytndkililyfnklykkikdnsildmryennkfidisgygsnisingdvyiystnrnqf 61
```

```
Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFS 88
           ||||||||||||||||||||||||||||
Db      62 giysskpsevniaqnndiiyngryqnfs 89


RESULT  15
AAW68399
ID   AAW68399 standard; Protein; 448 AA.
XX
AC   AAW68399;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type F toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US15394.
XX
PR   28-AUG-1996;   96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS,  Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment
CC   of Clostridium botulinum (202F strain) type F neurotoxin, encoded
CC   by a DNA sequence (see AAV30593) in plasmid pETHisb.  This vector
CC   can be used to express BotC soluble C fragment in Escherichia
CC   coli host cells, with the recombinant C fragment being purified on
CC   an affinity column.  The invention relates to recombinant proteins
CC   derived from C. botulinum toxins, especially type B and type E
CC   toxins.  Methods are provided which allow for the isolation of
CC   soluble recombinant proteins free of significant endotoxin
CC   contamination.  Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells.  The
CC   recombinant toxins are used as immunogens for the production of
CC   vaccines and antitoxins that are useful in the treatment of humans
CC   and animals at risk of intoxication with clostridial toxin.
XX
SQ   Sequence   448 AA;


  Query Match            18.1%;  Score 26;  DB 19;  Length 448;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-18;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 SILDMRYENNKFIDISGYGSNISING 48
           ||||||||||||||||||||||||||
Db      43 sildmryennkfidisgygsnising 68


Search completed: August 15, 2002, 11:12:26
Job time: 318 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:37 ; Search time 24.69 Seconds
(without alignments)
225.825 Million cell updates/sec

Title: US-08-981-087A-2
Perfect score: 144
Sequence: 1 SYTNDKILILYFNKLYKKIK.....LNYNKIIWTLQDTAGNNQKL 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	26	18.1	1274	1	BXF_CLOBO	P30996 clostridium
2	11	7.6	1250	1	BXE_CLOBO	Q00496 clostridium
3	11	7.6	1250	1	BXE_CLOBU	P30995 clostridium
4	9	6.2	1295	1	BXA2_CLOBO	Q45894 clostridium
5	8	5.6	1290	1	BXB_CLOBO	P10844 clostridium
6	8	5.6	1295	1	BXA1_CLOBO	P10845 clostridium
7	7	4.9	449	1	MURF_RICPR	O05953 rickettsia
8	7	4.9	501	1	VGLC_HSVMB	P22650 marek's dis
9	7	4.9	501	1	VGLC_HSVMD	P33500 marek's dis
10	7	4.9	501	1	VGLC_HSVMM	P22651 marek's dis
11	7	4.9	505	1	VGLC_HSVMG	P10681 marek's dis
12	7	4.9	511	1	RRB1_YEAST	Q04225 saccharomyc
13	7	4.9	2292	1	POLG_EMCVB	P17593 encephalomy
14	7	4.9	2292	1	POLG_EMCVD	P17594 encephalomy
15	6	4.2	79	1	CATR_HUMAN	Q13166 homo sapien
16	6	4.2	112	1	YF88_METJA	Q58983 methanococc
17	6	4.2	117	1	RK20_ASTLO	P34769 astasia lon
18	6	4.2	134	1	GILZ_HUMAN	Q99576 homo sapien
19	6	4.2	134	1	GILZ_RAT	Q9eqz1 rattus norv
20	6	4.2	140	1	LYC_ANOGA	Q17005 anopheles g
21	6	4.2	153	1	COXZ_COWRU	Q04302 cowdria rum
22	6	4.2	171	1	BB19_RABIT	Q05005 oryctolagus
23	6	4.2	176	1	CFTR_MACMU	Q00553 macaca mula
24	6	4.2	180	1	CFTR_CAVPO	Q00552 cavia porce
25	6	4.2	180	1	Y426_METJA	Q57869 methanococc
26	6	4.2	193	1	Y262_HELPY	O25042 helicobacte
27	6	4.2	194	1	YPBH_BACSU	P50734 bacillus su
28	6	4.2	196	1	TSA1_CANAL	Q9y7f0 candida alb
29	6	4.2	204	1	VATD_AERPE	Q9yf38 aeropyrum p
30	6	4.2	205	1	CYOC_BUCAI	P57542 buchnera ap
31	6	4.2	211	1	YP78_CAEEL	Q09224 caenorhabdi
32	6	4.2	214	1	CYB_AGKCO	P92845 agkistrodon
33	6	4.2	214	1	CYB_AGKHA	P92852 agkistrodon
34	6	4.2	214	1	CYB_BOTSC	P92849 bothriechis
35	6	4.2	217	1	EXPI_ERWCA	P33882 erwinia car
36	6	4.2	217	1	YE9H_SCHPO	O13777 schizosacch
37	6	4.2	218	1	Y010_MYCGE	P47256 mycoplasma
38	6	4.2	226	1	TDX1_CAEEL	Q21824 caenorhabdi
39	6	4.2	243	1	NGF_BUNMU	P34128 bungarus mu
40	6	4.2	245	1	PLSC_SALTY	P26974 salmonella
41	6	4.2	247	1	Y276_BUCAI	P57364 buchnera ap
42	6	4.2	256	1	HYPA_HYPLI	P35587 hypoderma l
43	6	4.2	257	1	TPI1_GIALA	P36186 giardia lam
44	6	4.2	257	1	TPI2_GIALA	P36187 giardia lam
45	6	4.2	278	1	NIFH_METVO	P06119 methanococc

ALIGNMENTS

```
RESULT   1
BXF_CLOBO
ID   BXF_CLOBO      STANDARD;      PRT;  1274 AA.
AC   P30996;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE   (Bontoxilysin F).
GN   BOTF.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 23387;
RX   MEDLINE=93012902; PubMed=1398040;
RA   East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA   Roberts T.A., Thompson D.E.;
RT   "Sequence of the gene encoding type F neurotoxin of Clostridium
RT   botulinum.";
RL   FEMS Microbiol. Lett. 75:225-230(1992).
RN   [2]
RP   SEQUENCE OF 1-64 FROM N.A.
RC   STRAIN=HOBBS FT10;
RX   MEDLINE=94297488; PubMed=7764998;
RA   East A.K., Collins M.D.;
RT   "Conserved structure of genes encoding components of botulinum
RT   neurotoxin complex M and the sequence of the gene coding for the
RT   nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL   Curr. Microbiol. 29:69-77(1994).
RN   [3]
RP   SEQUENCE OF 634-1002 FROM N.A.
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94230352; PubMed=8175689;
RA   Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA   Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT   "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT   F botulinal neurotoxins and tetanus toxin.";
RL   J. Biol. Chem. 269:12764-12772(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC       BOND OF SYNAPTOBREVINS-1 AND -2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
```

```
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDöm; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;


  Query Match             18.1%;  Score 26;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-19;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 SILDMRYENNKFIDISGYGSNISING 48
           ||||||||||||||||||||||||||
Db     869 SILDMRYENNKFIDISGYGSNISING 894


RESULT   2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BELUGA;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; A60027; A60027.
```

```
DR   PIR; B35294; B35294.
DR   PIR; JH0257; JH0257.
DR   PIR; S08575; S08575.
DR   PIR; S18111; S18111.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;


  Query Match             7.6%;  Score 11;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||
Db     913 NFSISFWVRIP 923


RESULT   3
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, AND ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   PIR; S16145; S16145.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;


  Query Match             7.6%;  Score 11;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||
Db     913 NFSISFWVRIP 923


RESULT   4
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
```

```
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;


  Query Match             6.2%;  Score 9;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.17;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     103 NLNNEYTII 111
           |||||||||
Db     956 NLNNEYTII 964


RESULT   5
BXB_CLOBO
ID   BXB_CLOBO      STANDARD;      PRT;  1290 AA.
AC   P10844; P10843;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type B precursor (EC 3.4.24.69) (BoNT/B)
DE   (Bontoxilysin B).
GN   BOTB.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92384550; PubMed=1514783;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
RA   Minton N.P.;
RT   "Molecular cloning of the Clostridium botulinum structural gene
RT   encoding the type B neurotoxin and determination of its entire
RT   nucleotide sequence.";
RL   Appl. Environ. Microbiol. 58:2345-2354(1992).
RN   [2]
RP   SEQUENCE OF 35-245 FROM N.A.
RC   STRAIN=NCTC 7273;
RA   Szabo E.A., Pemberton J.M., Desmarchelier P.M.;
RL   Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 633-993 FROM N.A.
RC   STRAIN=NCTC 7273;
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
RN   [4]
RP   SEQUENCE OF 1-44 AND 441-466.
RC   STRAIN=657;
RX   MEDLINE=89000987; PubMed=3139097;
RA   Dasgupta B.R., Datta A.;
RT   "Botulinum neurotoxin type B (strain 657): partial sequence and
RT   similarity with tetanus toxin.";
RL   Biochimie 70:811-817(1988).
RN   [5]
RP   SEQUENCE OF 1-16 AND 441-458.
RC   STRAIN=OKRA;
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [6]
RP   IDENTIFICATION AS ZINC-PROTEASE.
RX   MEDLINE=93054694; PubMed=1429690;
RA   Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.;
RT   "Botulinum neurotoxins are zinc proteins.";
RL   J. Biol. Chem. 267:23479-23483(1992).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
```

```
RX   MEDLINE=93063293; PubMed=1331807;
RA   Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P.,
RA   Dasgupta B.R., Montecucco C.;
RT   "Tetanus and botulinum-B neurotoxins block neurotransmitter release
RT   by proteolytic cleavage of synaptobrevin.";
RL   Nature 359:832-835(1992).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-|-PHE-77 BOND OF
CC       SYNAPTOBREVIN-2.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81186; AAA23211.1; -.
DR   EMBL; Z11934; CAA77991.1; -.
DR   EMBL; X70817; CAA50148.1; -.
DR   PIR; S07128; S07128.
DR   PIR; S07155; S07155.
DR   PIR; S08562; S08562.
DR   PIR; S08573; S08573.
DR   PIR; S08574; S08574.
DR   PIR; A48940; A48940.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    440       BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
FT   CHAIN       441   1290       BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    436    445       INTERCHAIN (PROBABLE).
FT   CONFLICT     29     29       T -> M (IN REF. 4).
FT   CONFLICT    217    217       R -> G (IN REF. 2).
FT   CONFLICT    224    224       A -> S (IN REF. 2).
FT   CONFLICT    463    463       S -> R (IN REF. 4).
SQ   SEQUENCE   1290 AA;  150670 MW;  D21746E2C024DF43 CRC64;


  Query Match             5.6%;  Score 8;  DB 1;  Length 1290;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     957 NNSGWKIS 964
```

```
RESULT   6
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P18639; P01561;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=HALL;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
```

```
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX(R) (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PIR; S09492; S09492.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;


  Query Match             5.6%;  Score 8;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     104 LNNEYTII 111
           ||||||||
Db     957 LNNEYTII 964


RESULT   7
MURF_RICPR
ID   MURF_RICPR     STANDARD;      PRT;   449 AA.
AC   O05953;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-
DE   alanyl-D-alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide
DE   synthetase) (D-alanyl-D-alanine-adding enzyme).
GN   MURF OR RP596.
OS   Rickettsia prowazekii.
```

```
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE OF 1-96 FROM N.A.
RC   STRAIN=MADRID E;
RX   MEDLINE=97419517; PubMed=9274032;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
CC   -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC       STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC       PRECURSOR OF MUREIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC       glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC       + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC       carboxy-L-lysyl-D-alanyl-D-alanine.
CC   -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235272; CAA15040.1; -.
DR   EMBL; Y11783; CAA72472.1; -.
DR   HSSP; P11880; 1GG4.
DR   InterPro; IPR000713; Mur_ligase.
DR   InterPro; IPR004101; Mur_ligase_C.
DR   Pfam; PF01225; Mur_ligase; 1.
DR   Pfam; PF02875; Mur_ligase_C; 1.
KW   Peptidoglycan synthesis; Cell wall; Cell division; Ligase;
KW   ATP-binding; Complete proteome.
FT   NP_BIND     106    112       ATP (POTENTIAL).
SQ   SEQUENCE   449 AA;  50672 MW;  3FEB8468F825BFDA CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 449;
  Best Local Similarity   100.0%;  Pred. No. 8.9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      31 NNKFIDI 37
           |||||||
Db     268 NNKFIDI 274


RESULT   8
VGLC_HSVMB
ID   VGLC_HSVMB     STANDARD;      PRT;   501 AA.
AC   P22650;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)
DE   (GA).
GN   GA.
OS   Marek's disease herpesvirus (strain bc-1) (MDHV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10387;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90142542; PubMed=2559540;
RA   Ihara T., Kato A., Ueda S., Ishihama A., Hirai K.;
RT   "Comparison of the sequence of the secretory glycoprotein A (gA) gene
RT   in Md5 and BC-1 strains of Marek's disease virus type 1.";
RL   Virus Genes 3:127-140(1989).
CC   -!- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF
CC       MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE
CC       IMMUNOSUPPRESSION THAT OCCURS AFTER MDHV INFECTION.
CC   -!- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT
CC       OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY
CC       OTHER INTERACTIONS.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90002; BAA14054.1; ALT_SEQ.
DR   PIR; JS0388; VGBEMA.
DR   InterPro; IPR001654; Marek_A.
DR   Pfam; PF02124; Marek_A; 1.
DR   PRINTS; PR00675; MAREKSGPA.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    501       SECRETORY GLYCOPROTEIN GP57-65.
FT   TRANSMEM    466    492       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD     46     46       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    100    100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    120    120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212    212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    354    354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400    400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    429    429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    493    493       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   501 AA;  56134 MW;  DE71E952331024B0 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 SNISING 48
           |||||||
Db      45 SNISING 51


RESULT   9
VGLC_HSVMD
ID   VGLC_HSVMD     STANDARD;      PRT;   501 AA.
AC   P33500;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)
DE   (GA).
GN   GA.
OS   Marek's disease herpesvirus (strain RB-1B) (MDHV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=33707;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RX   MEDLINE=89269090; PubMed=2543160;
RA   Binns M.M., Ross N.L.J.;
RT   "Nucleotide sequence of the Marek's disease virus (MDV) RB-1B A
RT   antigen gene and the identification of the MDV A antigen as the
RT   herpes simplex virus-1 glycoprotein C homologue.";
RL   Virus Res. 12:371-382(1989).
CC   -!- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF
CC       MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE
CC       IMMUNOSUPPRESSION THAT OCCURS AFTER MDHV INFECTION.
CC   -!- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT
CC       OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY
CC       OTHER INTERACTIONS.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
DR   PIR; A60005; A60005.
DR   InterPro; IPR001654; Marek_A.
DR   Pfam; PF02124; Marek_A; 1.
DR   PRINTS; PR00675; MAREKSGPA.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    501       SECRETORY GLYCOPROTEIN GP57-65.
FT   TRANSMEM    466   492       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD     46    46       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91    91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    100   100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    120   120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212   212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    354   354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400   400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    429   429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    493   493       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   501 AA;  56104 MW;  438473B1DD79340A CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 SNISING 48
           |||||||
Db      45 SNISING 51


RESULT  10
VGLC_HSVMM
ID   VGLC_HSVMM     STANDARD;      PRT;   501 AA.
AC   P22651;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)
DE   (GA).
GN   GA.
OS   Marek's disease herpesvirus (strain Md5) (MDHV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10389;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90142542; PubMed=2559540;
RA   Ihara T., Kato A., Ueda S., Ishihama A., Hirai K.;
RT   "Comparison of the sequence of the secretory glycoprotein A (gA) gene
RT   in Md5 and BC-1 strains of Marek's disease virus type 1.";
RL   Virus Genes 3:127-140(1989).
CC   -!- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF
CC       MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE
CC       IMMUNOSUPPRESSION THAT OCCURS AFTER MDHV INFECTION.
CC   -!- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT
CC       OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY
CC       OTHER INTERACTIONS.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90001; BAA14052.1; -.
DR   PIR; A22818; VGBEMB.
DR   InterPro; IPR001654; Marek_A.
DR   Pfam; PF02124; Marek_A; 1.
DR   PRINTS; PR00675; MAREKSGPA.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL       1     27       POTENTIAL.
FT   CHAIN       28    501       SECRETORY GLYCOPROTEIN GP57-65.
FT   TRANSMEM    466   492       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD     46    46       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91    91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    100   100       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    120   120       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    212   212       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    354   354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    400   400       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    429   429       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    493   493       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   501 AA;  56088 MW;  4393C56AA779340A CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 501;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 SNISING 48
           |||||||
Db      45 SNISING 51


RESULT  11
VGLC_HSVMG
ID   VGLC_HSVMG     STANDARD;      PRT;   505 AA.
AC   P10681;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Secretory glycoprotein GP57-65 precursor (A antigen) (Glycoprotein A)
DE   (GA).
GN   GA.
OS   Marek's disease herpesvirus (strain GA) (MDHV).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10388;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88230597; PubMed=2836620;
RA   Coussens P.M., Velicer L.F.;
RT   "Structure and complete nucleotide sequence of the Marek's disease
RT   herpesvirus gp57-65 gene.";
RL   J. Virol. 62:2373-2379(1988).
CC   -!- FUNCTION: MAY PLAY AN IMMUNOEVASIVE ROLE IN THE PATHOGENESIS OF
CC       MAREK'S DISEASE. IT IS A CANDIDATE FOR CAUSING THE EARLY-STAGE
CC       IMMUNOSUPPRESSION THAT OCCURS AFTER MDHV INFECTION.
CC   -!- SUBCELLULAR LOCATION: PREDOMINANTLY SECRETED, BUT A SMALL AMOUNT
CC       OF MATURE GP57-65 IS ANCHORED IN THE PLASMA MEMBRANE OR HELD BY
CC       OTHER INTERACTIONS.
CC   -!- SIMILARITY: BELONGS TO THE HERPESVIRUSES GLYCOPROTEIN C FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; M20001; AAA46114.1; -.
DR   PIR; A28843; VGBEMH.
DR   InterPro; IPR001654; Marek_A.
DR   Pfam; PF02124; Marek_A; 1.
DR   PRINTS; PR00675; MAREKSGPA.
KW   Glycoprotein; Transmembrane; Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   CHAIN        28    505       SECRETORY GLYCOPROTEIN GP57-65.
FT   TRANSMEM    465    491       MEMBRANE ANCHOR (POTENTIAL).
FT   CARBOHYD     45     45       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     90     90       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    119    119       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    211    211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    399    399       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    428    428       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   505 AA;  56809 MW;  D06D75D7D9C6666D CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 505;
  Best Local Similarity   100.0%;  Pred. No. 9.8;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      42 SNISING 48
           |||||||
Db      44 SNISING 50


RESULT  12
RRB1_YEAST
ID   RRB1_YEAST     STANDARD;      PRT;   511 AA.
AC   Q04225;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ribosome assembly protein RRB1.
GN   RRB1 OR YMR131C OR YM9553.07C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Badcock K., Churcher C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=21585391; PubMed=11728313;
RA   Schaper S., Fromont-Racine M., Linder P., de la Cruz J., Namane A.,
RA   Yaniv M.;
RT   "A yeast homolog of chromatin assembly factor 1 is involved in early
RT   ribosome assembly.";
RL   Curr. Biol. 11:1885-1890(2001).
CC   -!- FUNCTION: Involved in regulation of L3 expression and stability
CC       and plays a role in early 60S ribosomal subunit assembly. May be
CC       required for proper assembly of preribosomal particles during
CC       early ribosome biogenesis, presumably by targeting L3 onto the 35S
CC       precursor rRNA.
CC   -!- SUBUNIT: Associates with ribosomal protein L3.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z48622; CAA88556.1; -.
DR   SGD; S0004738; RRB1.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 4.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 4.
DR   PROSITE; PS00678; WD_REPEATS_1; 1.
DR   PROSITE; PS50082; WD_REPEATS_2; 2.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Ribosome biogenesis; rRNA processing; Nuclear protein; Repeat;
KW   WD repeat.
FT   REPEAT      319    359       WD 1.
FT   REPEAT      364    404       WD 2.
FT   REPEAT      415    455       WD 3.
FT   REPEAT      477    510       WD 4.
SQ   SEQUENCE   511 AA;  57261 MW;  1D18CE3C60BAFF30 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 511;
  Best Local Similarity   100.0%;  Pred. No. 9.9;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      47 NGDVYIY 53
           |||||||
Db     235 NGDVYIY 241


RESULT  13
POLG_EMCVB
ID   POLG_EMCVB     STANDARD;      PRT;  2292 AA.
AC   P17593;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE   P2A TO P2C, P3A; Genome-linked protein VPG; Picornain 3C
DE   (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
DE   (EC 2.7.7.48)].
OS   Encephalomyocarditis virus (strain emc-b nondiabetogenic).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Cardiovirus.
OX   NCBI_TaxID=12105;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89243189; PubMed=2541543;
RA   Bae Y.S., Eun H.M., Yoon J.W.;
RT   "Genomic differences between the diabetogenic and nondiabetogenic
RT   variants of encephalomyocarditis virus.";
RL   Virology 170:282-287(1989).
CC   -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC       Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22457; AAA43033.1; ALT_SEQ.
DR   PIR; B31473; GNNYEB.
DR   HSSP; P12296; 2MEV.
DR   MEROPS; C03.009; -.
DR   MEROPS; U29.001; -.
DR   InterPro; IPR000605; RNA_helicase.
DR   InterPro; IPR001205; RNA_pol_P3D.
DR   InterPro; IPR001676; Rhv.
DR   Pfam; PF00073; rhv; 3.
```

```
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT   PROPEP          1     67       LEADER PEPTIDE.
FT   CHAIN          68    137       COAT PROTEIN VP4 (RHO).
FT   CHAIN         138    393       COAT PROTEIN VP2 (BETA).
FT   CHAIN         394    624       COAT PROTEIN VP3 (GAMMA).
FT   CHAIN         625    901       COAT PROTEIN VP1 (ALPHA).
FT   CHAIN         902   1058       CORE PROTEIN P2A (G).
FT   CHAIN        1059   1194       CORE PROTEIN P2B (I).
FT   CHAIN        1195   1519       CORE PROTEIN P2C (F).
FT   CHAIN        1520   1607       CORE PROTEIN P3A.
FT   CHAIN        1608   1627       GENOME-LINKED PROTEIN VPG (H).
FT   CHAIN        1628   1832       PICORNAIN 3C (P22).
FT   CHAIN        1833   2292       RNA-DIRECTED RNA POLYMERASE P3D (E).
FT   LIPID          68     68       MYRISTATE (BY SIMILARITY).
FT   ACT_SITE     1786   1786       PROTEASE (POTENTIAL).
FT   ACT_SITE     1804   1804       PROTEASE (POTENTIAL).
SQ   SEQUENCE   2292 AA;  255495 MW;  8540D0EB1437E8D4 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 2292;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 LQDTAGN 140
           |||||||
Db     152 LQDTAGN 158


RESULT  14
POLG_EMCVD
ID   POLG_EMCVD     STANDARD;      PRT;  2292 AA.
AC   P17594;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Genome polyprotein [Contains: Coat proteins VP1 TO VP4; Core proteins
DE   P2A TO P2C, P3A; Genome-linked protein VPG; Picornain 3C
DE   (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase P3D
DE   (EC 2.7.7.48)].
OS   Encephalomyocarditis virus (strain emc-d diabetogenic).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;
OC   Cardiovirus.
OX   NCBI_TaxID=12106;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89243189; PubMed=2541543;
RA   Bae Y.S., Eun H.M., Yoon J.W.;
RT   "Genomic differences between the diabetogenic and nondiabetogenic
RT   variants of encephalomyocarditis virus.";
RL   Virology 170:282-287(1989).
CC   -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC       Q/G SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC   -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAHEDRAL UNITS,
CC       EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC       VP3, AND VP4.
CC   -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M22458; AAA43034.1; -.
DR   PIR; A31473; GNNYED.
DR   HSSP; P12296; 2MEV.
DR   MEROPS; C03.009; -.
DR   MEROPS; U29.001; -.
DR   InterPro; IPR000605; RNA_helicase.
DR   InterPro; IPR001205; RNA_pol_P3D.
DR   InterPro; IPR001676; Rhv.
DR   Pfam; PF00073; rhv; 3.
DR   Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR   Pfam; PF00910; RNA_helicase; 1.
KW   Polyprotein; Coat protein; Core protein; Transferase;
KW   RNA-directed RNA polymerase; Hydrolase; Thiol protease; Myristate.
FT   PROPEP          1     67       LEADER PEPTIDE.
FT   CHAIN          68    137       COAT PROTEIN VP4 (RHO).
FT   CHAIN         138    393       COAT PROTEIN VP2 (BETA).
FT   CHAIN         394    624       COAT PROTEIN VP3 (GAMMA).
FT   CHAIN         625    901       COAT PROTEIN VP1 (ALPHA).
FT   CHAIN         902   1058       CORE PROTEIN P2A (G).
FT   CHAIN        1059   1194       CORE PROTEIN P2B (I).
FT   CHAIN        1195   1519       CORE PROTEIN P2C (F).
FT   CHAIN        1520   1607       CORE PROTEIN P3A.
FT   CHAIN        1608   1627       GENOME-LINKED PROTEIN VPG (H).
FT   CHAIN        1628   1832       PICORNAIN 3C (P22).
FT   CHAIN        1833   2292       RNA-DIRECTED RNA POLYMERASE P3D (E).
FT   LIPID          68     68       MYRISTATE (BY SIMILARITY).
FT   ACT_SITE     1786   1786       PROTEASE (POTENTIAL).
FT   ACT_SITE     1804   1804       PROTEASE (POTENTIAL).
SQ   SEQUENCE   2292 AA;  255426 MW;  F2B0627B0F444107 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 2292;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     134 LQDTAGN 140
           |||||||
Db     152 LQDTAGN 158


RESULT  15
CATR_HUMAN
ID   CATR_HUMAN     STANDARD;      PRT;    79 AA.
AC   Q13166;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   CATR tumorigenic conversion 1 protein (CATR1.3).
GN   CATR1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Carcinoma;
RX   MEDLINE=95327656; PubMed=7604004;
RA   Li D., Noyes I., Shuler C., Milo G.E.;
RT   "Cloning and sequencing of CATR1.3, a human gene associated with
RT   tumorigenic conversion.";
RL   Proc. Natl. Acad. Sci. U.S.A. 92:6409-6413(1995).
CC   -!- DEVELOPMENTAL STAGE: ASSOCIATED WITH TUMORIGENIC CONVERSION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U25433; -; NOT_ANNOTATED_CDS.
DR   MIM; 600676; -.
SQ   SEQUENCE   79 AA;  9224 MW;  BC3667C059114CF3 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 79;
```

Best Local Similarity 100.0%; Pred. No. 24;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      74 QNNDII 79
           ||||||
Db      15 QNNDII 20
```

Search completed: August 15, 2002, 11:24:38
Job time: 685 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:26 ; Search time 96.53 Seconds
(without alignments)
165.696 Million cell updates/sec

Title: US-08-981-087A-3
Perfect score: 144
Sequence: 1 VFNYTQMISISDYINKWIFV.....ITQNSNFLNINQQRGVYQKP 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144	100.0	144	18	AAW09016	Immunogenic type F
2	144	100.0	431	18	AAW09014	Immunogenic type F
3	144	100.0	432	21	AAY77138	Synthetic botulinu
4	144	100.0	432	22	AAB04096	Botulism toxin hea
5	144	100.0	432	22	AAB04103	Botulism toxin hea
6	144	100.0	645	22	AAE07894	Modified clostridi
7	144	100.0	685	22	AAE07893	Modified clostridi
8	144	100.0	862	22	AAE07890	Modified clostridi
9	144	100.0	887	22	AAE07892	Modified clostridi
10	144	100.0	1032	22	AAE07901	C. botulinum C2 tr
11	144	100.0	1059	21	AAY93309	A manganese supero
12	144	100.0	1084	21	AAY93312	A manganese supero
13	144	100.0	1092	22	AAE07900	C. botulinum C2 tr
14	69	47.9	660	22	AAE07898	Modified clostridi
15	27	18.8	448	19	AAW68399	Clostridium botuli
16	15	10.4	419	22	AAB04095	Botulism toxin hea
17	15	10.4	449	21	AAY77137	Synthetic botulinu
18	15	10.4	449	22	AAB04094	Botulism toxin hea
19	15	10.4	451	19	AAW68395	Clostridium botuli
20	15	10.4	452	19	AAW68396	Clostridium botuli
21	11	7.6	233	21	AAY77143	Botulinum neurotox
22	11	7.6	382	21	AAB36303	BoNT/A prototoxin
23	11	7.6	415	22	AAB04083	Botulism toxin C f
24	11	7.6	432	21	AAY77142	Native botulinum n
25	11	7.6	434	22	AAB04089	Botulism toxin hea
26	11	7.6	435	22	AAB04090	Botulism toxin hea
27	11	7.6	437	22	AAB04088	Botulism toxin hea
28	11	7.6	438	17	AAR95008	Type A neurotoxin
29	11	7.6	438	19	AAW68389	Clostridium botuli
30	11	7.6	438	21	AAY77134	Synthetic botulinu
31	11	7.6	445	19	AAW68391	Clostridium botuli
32	11	7.6	462	17	AAR95009	Type A neurotoxin
33	11	7.6	462	19	AAW68390	Clostridium botuli
34	11	7.6	837	21	AAY77140	Native botulinum n
35	11	7.6	847	22	AAB04081	Botulism toxin hea
36	11	7.6	1067	21	AAY93307	A manganese supero
37	11	7.6	1092	21	AAY93310	A manganese supero
38	11	7.6	1296	17	AAR95010	C. botulinum type
39	8	5.6	449	21	AAY77139	Synthetic botulinu
40	8	5.6	449	22	AAB04167	Botulism toxin hea
41	8	5.6	473	19	AAW68400	Clostridium botuli
42	7	4.9	20	16	AAR64982	MMLV p15E C1q bind
43	7	4.9	196	15	AAR43986	p15e Protein. Hum
44	7	4.9	196	15	AAR55343	Sequence of envelo
45	7	4.9	206	21	AAY77144	Botulinum neurotox

ALIGNMENTS

RESULT 1
AAW09016
ID AAW09016 standard; Protein; 144 AA.
XX
AC AAW09016;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin polypeptide (aa992-1135).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 18-19; 37pp; English.
XX
CC Novel polypeptides (AAW09014-17) respectively comprise amino acids

CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a
CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and
CC HN epitopes necessary for metalloprotease activity and toxin
CC internalisation. They are free of botulinum toxin activity but can
CC induce protective immunity to a type F botulinum toxin, making them
CC useful for vaccine prodn. Recombinant polypeptides can be
CC produced in transformed host cells, esp. as fusion proteins, e.g.
CC with maltose binding protein to facilitate purification.
XX
SQ Sequence 144 AA;

Query Match 100.0%; Score 144; DB 18; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.4e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 1 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 60

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 61 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 120

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 121 tdksitqnsnflninqqrgvyqkp 144

RESULT 2
AAW09014
ID AAW09014 standard; Protein; 431 AA.
XX
AC AAW09014;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX
CC A polypeptide (AAW09014) comprises the heavy chain (amino acids
CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be
CC produced using a synthetic gene (AAT48101) based on the natural
CC gene sequence (AAT48100) for the heavy chain. The polypeptides and
CC its fragments (see also AAW09015-17) lack the light chain and HN
CC epitopes necessary for metalloprotease activity and toxin
CC internalisation. They are free of botulinum toxin activity but can
CC induce protective immunity to a type F botulinum toxin, making them
CC useful for vaccine prodn. Recombinant polypeptides can be
CC produced in transformed host cells, esp. as fusion proteins, e.g.
CC with maltose binding protein to facilitate purification.
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 144; DB 18; Length 431;
Best Local Similarity 100.0%; Pred. No. 3.9e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 145 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 204

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 205 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 264

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 265 tdksitqnsnflninqqrgvyqkp 288

RESULT 3
AAY77138
ID AAY77138 standard; Protein; 432 AA.
XX
AC AAY77138;
XX
DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87216.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
PS Claim 27; Page -; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which

CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAY77134-Y77139 represent
CC synthetic BoNT Hc fragments used in the present invention. The DNA
CC encoding these sequences had been optimised for codon usage for
CC expression in yeast. Note: This sequence is not given in the
CC specification, but is decoded from the BoNTF Hc DNA sequence given on
CC pages 45-46.
XX
SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 21; Length 432;
Best Local Similarity 100.0%; Pred. No. 3.9e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 146 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 205

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 206 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 265

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 266 tdksitqnsnflninqqrgvyqkp 289

RESULT 4
AAB04096
ID AAB04096 standard; Protein; 432 AA.
XX
AC AAB04096;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54490.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Claim 3; Fig 9b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 3.9e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 146 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 205

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 206 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 265

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 266 tdksitqnsnflninqqrgvyqkp 289

RESULT 5
AAB04103
ID AAB04103 standard; Protein; 432 AA.
XX
AC AAB04103;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54499.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Disclosure; Fig 18b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 432 AA;

Query Match 100.0%; Score 144; DB 22; Length 432;
Best Local Similarity 100.0%; Pred. No. 3.9e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 146 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 205

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 206 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 265

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 266 tdksitqnsnflninqqrgvyqkp 289

RESULT 6
AAE07894
ID AAE07894 standard; Protein; 645 AA.
XX
AC AAE07894;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain fragment #1.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 44; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;

Query Match 100.0%; Score 144; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 5.7e-147;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
Db 359 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 418

Qy 61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
Db 419 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 478

Qy 121 TDKSITQNSNFLNINQQRGVYQKP 144
Db 479 tdksitqnsnflninqqrgvyqkp 502

RESULT 7
AAE07893
ID AAE07893 standard; Protein; 685 AA.
XX
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.

XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 43; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation peptide from influenza virus and a neuronal cell-specific
CC binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 685 AA;

```
  Query Match             100.0%;  Score 144;  DB 22;  Length 685;
  Best Local Similarity   100.0%;  Pred. No. 6e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 458

Qy      61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 518

Qy     121 TDKSITQNSNFLNINQQRGVYQKP 144
           ||||||||||||||||||||||||
Db     519 tdksitqnsnflninqqrgvyqkp 542
```

RESULT 8
AAE07890
ID AAE07890 standard; Protein; 862 AA.
XX
AC AAE07890;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 40; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate.
CC This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation domain from diphtheria neurotoxin and a neuronal cell-
CC specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 862 AA;

```
  Query Match             100.0%;  Score 144;  DB 22;  Length 862;
  Best Local Similarity   100.0%;  Pred. No. 7.5e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     576 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 635

Qy      61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     636 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 695

Qy     121 TDKSITQNSNFLNINQQRGVYQKP 144
           ||||||||||||||||||||||||
Db     696 tdksitqnsnflninqqrgvyqkp 719
```

RESULT 9
AAE07892
ID AAE07892 standard; Protein; 887 AA.
XX
AC AAE07892;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.

```
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Bacillus stearothermophilus.
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
OS   Chimeric - Synthetic.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 9; Page 42; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain-superoxide dismutase conjugate.
CC   This conjugate comprises a mitochondrial leader sequence from human
CC   Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC   linker that can be cleaved by thrombin, translocation domain from
CC   diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC   botulinum neurotoxin type F (BoNT/F).
XX
SQ   Sequence   887 AA;

  Query Match             100.0%;  Score 144;  DB 22;  Length 887;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 660

Qy      61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 720

Qy     121 TDKSITQNSNFLNINQQRGVYQKP 144
           ||||||||||||||||||||||||
Db     721 tdksitqnsnflninqqrgvyqkp 744
```

```
RESULT  10
AAE07901
ID   AAE07901 standard; Protein; 1032 AA.
XX
AC   AAE07901;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 48; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1032 AA;

  Query Match             100.0%;  Score 144;  DB 22;  Length 1032;
  Best Local Similarity   100.0%;  Pred. No. 8.9e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     746 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 805

Qy      61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     806 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 865

Qy     121 TDKSITQNSNFLNINQQRGVYQKP 144
           ||||||||||||||||||||||||
Db     866 tdksitqnsnflninqqrgvyqkp 889
```

```
RESULT  11
AAY93309
ID   AAY93309 standard; protein; 1059 AA.
XX
AC   AAY93309;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 48-51; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC   can be cleaved by thrombin, and a heavy chain derived from botulinum
CC   neurotoxin serotype F. The specification describes a composition for
CC   delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC   by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC   This component has a domain that binds to a neuronal cell and a
CC   domain that translocates the SOD of the composition into the neuronal
CC   cell. After translocation, the linker is cleaved to release the SOD.
CC   The composition is useful for treating neuronal diseases caused or
CC   augmented by oxidative stress, such as ischemic stroke, trauma,
CC   Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ   Sequence   1059 AA;


  Query Match           100.0%;  Score 144;  DB 21;  Length 1059;
  Best Local Similarity  100.0%;  Pred. No. 9.1e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 832

Qy     61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 892

Qy    121 TDKSITQNSNFLNINQQRGVYQKP 144
          ||||||||||||||||||||||||
Db    893 tdksitqnsnflninqqrgvyqkp 916
```

```
RESULT  12
AAY93312
ID   AAY93312 standard; protein; 1084 AA.
XX
AC   AAY93312;
XX
DT   04-SEP-2000  (first entry)
XX
DE   A manganese superoxide dismutase (Mn-SOD) construct.
XX
KW   Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW   neuronal cell targeting component; NCTC; neuronal disease;
KW   oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW   Huntington's disease; motor neurone disease;
KW   botulinum neurotoxin serotype F.
XX
OS   Synthetic.
OS   Homo sapiens.
OS   Bacillus stearothermophilus.
OS   Clostridium botulinum.
XX
PN   WO200028041-A1.
XX
PD   18-MAY-2000.
XX
PF   05-NOV-1999;   99WO-GB03699.
XX
PR   05-NOV-1998;   98GB-0024282.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Hallis B,  Silman N;
XX
DR   WPI; 2000-376553/32.
XX
PT   Novel composition, comprising superoxide dismutase linked by a
PT   cleavable linker to a neuronal cell targeting component useful for
PT   delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS   Disclosure; Page 57-60; 65pp; English.
XX
CC   The present sequence represents a construct of the invention, comprising
CC   a mitochondrial leader sequence from human manganese superoxide
CC   dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC   that can be cleaved by thrombin, and a heavy chain derived from
CC   botulinum neurotoxin serotype F. The specification describes a
CC   composition for delivery of SOD to neuronal cells. The composition
CC   comprises SOD linked, by a cleavable linker, to a neuronal cell
CC   targeting component (NCTC). This component has a domain that binds
CC   to a neuronal cell and a domain that translocates the SOD of the
CC   composition into the neuronal cell. After translocation, the linker
CC   is cleaved to release the SOD. The composition is useful for treating
CC   neuronal diseases caused or augmented by oxidative stress, such as
CC   ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC   motor neurone diseases.
XX
SQ   Sequence   1084 AA;


  Query Match           100.0%;  Score 144;  DB 21;  Length 1084;
  Best Local Similarity  100.0%;  Pred. No. 9.3e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 857

Qy     61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 917

Qy    121 TDKSITQNSNFLNINQQRGVYQKP 144
```

```
          ||||||||||||||||||||||||
Db    918 tdksitqnsnflninqqrgvyqkp 941


RESULT  13
AAE07900
ID   AAE07900 standard; Protein; 1092 AA.
XX
AC   AAE07900;
XX
DT   01-NOV-2001  (first entry)
XX
DE   C. botulinum C2 translocation domain with BoNT/F-binding domain #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 47; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is C. botulinum C2 enterotoxin translocation domain with botulinum
CC   neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC   the invention.
XX
SQ   Sequence   1092 AA;


  Query Match           100.0%;  Score 144;  DB 22;  Length 1092;
  Best Local Similarity  100.0%;  Pred. No. 9.4e-147;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    806 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 865

Qy     61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    866 vgcndtryvgiryfkvfdtelgkteietlysdepdpsilkdfwgnyllynkryyllnllr 925

Qy    121 TDKSITQNSNFLNINQQRGVYQKP 144
          ||||||||||||||||||||||||
Db    926 tdksitqnsnflninqqrgvyqkp 949


RESULT  14
AAE07898
ID   AAE07898 standard; Protein; 660 AA.
XX
AC   AAE07898;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #5.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; tetanus neurotoxin; TeNT;
KW   botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium tetani.
OS   Chimeric - Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
XX
PS   Example 2; Page 46; 50pp; English.
XX
CC   The invention relates to a non toxic polypeptide, for delivery of a
CC   therapeutic agent to a neuronal cell, which comprises a binding domain
CC   (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC   as Hc) that binds to the neuronal cell and a translocation domain (amino
CC   terminal half of HC, designated as HN), that translocates the therapeutic
CC   agent into the neuronal cell, where the translocation domain is not a HN
CC   domain of a clostridial neurotoxin and is not a fragment or derivative of
CC   a HN domain of a clostridial toxin. Polypeptides of the invention are
CC   useful for the treatment of a disease state associated with neuronal
CC   cells. The polypeptide constructs are useful for delivering therapeutic
CC   substances to neuronal cells. They are useful to treat disorders of the
CC   CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC   and infection. They are also useful in gene therapy. The present sequence
CC   is modified clostridial heavy chain fragment. This sequence is
CC   constructed by fusing the binding domain which is a hybrid of botulinum
CC   neurotoxin type F (BoNT/F) and tetanus neurotoxin (TeNT) domain II with
CC   translocation domain of diphtheria neurotoxin.
XX
SQ   Sequence   660 AA;


  Query Match           47.9%;  Score 69;  DB 22;  Length 660;
  Best Local Similarity  100.0%;  Pred. No. 7.1e-66;
  Matches   69;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    359 vfnytqmisisdyinkwifvtitnnrlgnsriyingnlideksisnlgdihvsdnilfki 418
```

```
Qy      61 VGCNDTRYV 69
           |||||||||
Db     419 vgcndtryv 427


RESULT  15
AAW68399
ID   AAW68399 standard; Protein; 448 AA.
XX
AC   AAW68399;
XX
DT   07-DEC-1998  (first entry)
XX
DE   Clostridium botulinum type F toxin C fragment.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW   botulism; BotF.
XX
OS   Clostridium botulinum serotype F strain 202F (ATCC 23387).
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..21
FT                   /note= "N-terminal His tag"
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997;   97WO-US15394.
XX
PR   28-AUG-1996;   96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS,  Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   N-PSDB; AAV30593.
XX
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin  - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
XX
PS   Example 48; Page 364-365; 428pp; English.
XX
CC   This is the amino acid sequence of the histidine-tagged C fragment
CC   of Clostridium botulinum (202F strain) type F neurotoxin, encoded
CC   by a DNA sequence (see AAV30593) in plasmid pETHisb.  This vector
CC   can be used to express BotC soluble C fragment in Escherichia
CC   coli host cells, with the recombinant C fragment being purified on
CC   an affinity column.  The invention relates to recombinant proteins
CC   derived from C. botulinum toxins, especially type B and type E
CC   toxins.  Methods are provided which allow for the isolation of
CC   soluble recombinant proteins free of significant endotoxin
CC   contamination.  Preferred hosts for production of recombinant
CC   proteins are E. coli, insect cells and yeast cells.  The
CC   recombinant toxins are used as immunogens for the production of
CC   vaccines and antitoxins that are useful in the treatment of humans
CC   and animals at risk of intoxication with clostridial toxin.
XX
SQ   Sequence   448 AA;


  Query Match             18.8%;  Score 27;  DB 19;  Length 448;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-20;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 YINKWIFVTITNNRLGNSRIYINGNLI 39
           |||||||||||||||||||||||||||
Db     180 yinkwifvtitnnrlgnsriyingnli 206
```

Search completed: August 15, 2002, 11:12:27
Job time: 319 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:06 ; Search time 47.36 Seconds
(without alignments)
292.164 Million cell updates/sec

Title: US-08-981-087A-3
Perfect score: 144
Sequence: 1 VFNYTQMISISDYINKWIFV.....ITQNSNFLNINQQRGVYQKP 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	18.8	1274	2	I40813	neurotoxin type F
2	22	15.3	1268	2	S33411	botulinum neurotox
3	15	10.4	1251	2	JH0256	botulinum neurotox
4	15	10.4	1252	2	S21178	botulinum neurotox
5	11	7.6	1296	1	BTCLAB	bontoxilysin (EC 3
6	11	7.6	1296	2	I40645	botulinum neurotox
7	9	6.2	366	2	S48110	neurotoxin type F
8	8	5.6	540	2	B87350	hypothetical prote
9	8	5.6	1297	2	S39791	neurotoxin - Clost
10	7	4.9	156	2	S20990	myosin regulatory
11	7	4.9	200	2	A81295	probable membrane
12	7	4.9	202	2	S20992	myosin regulatory
13	7	4.9	204	2	S22715	myosin regulatory
14	7	4.9	226	2	AI2668	hypothetical prote
15	7	4.9	231	2	S28494	dtxA protein - Clo
16	7	4.9	243	2	A91044	hypothetical prote
17	7	4.9	243	2	D85888	hypothetical prote
18	7	4.9	265	2	AG0814	cob(I)alamin adeno
19	7	4.9	267	2	G97450	ABC transporter, A
20	7	4.9	267	2	F81029	type II restrictio
21	7	4.9	267	2	B65021	cob(I)alamin adeno
22	7	4.9	307	2	C83188	probable ATP-bindi
23	7	4.9	358	2	S07594	hypothetical prote
24	7	4.9	379	2	B69344	hypothetical prote
25	7	4.9	432	2	D64512	hypothetical prote
26	7	4.9	446	2	T31644	hypothetical prote
27	7	4.9	461	2	B97305	probable cation ef
28	7	4.9	508	2	E69525	hypothetical prote
29	7	4.9	608	2	E71859	phosphogluconate d
30	7	4.9	608	2	D64657	phosphogluconate d
31	7	4.9	665	1	VCVWEM	env polyprotein -
32	7	4.9	814	2	B96630	Phosphatidylinosit
33	7	4.9	944	2	D82926	hypothetical prote
34	7	4.9	1023	2	T26261	hypothetical prote
35	7	4.9	1035	2	I58409	integrin alpha-9 c
36	7	4.9	1270	2	T21269	hypothetical prote
37	7	4.9	1276	2	S11455	botulinum neurotox
38	7	4.9	1291	1	A48940	bontoxilysin (EC 3
39	7	4.9	1291	2	I40631	non-proteolytic bo
40	7	4.9	1651	2	JC1340	outer membrane pro
41	7	4.9	1655	2	E97835	hypothetical prote
42	6	4.2	30	2	G95031	hypothetical prote
43	6	4.2	35	2	B86327	protein F18O14.13
44	6	4.2	48	1	EWAS8	H+-transporting tw
45	6	4.2	48	1	EWAS8M	H+-transporting tw

ALIGNMENTS

RESULT 1
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson,
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 18.8%; Score 27; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 4.9e-20;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 YINKWIFVTITNNRLGNSRIYINGNLI 39
||||||||||||||||||||||||||
Db 1006 YINKWIFVTITNNRLGNSRIYINGNLI 1032

RESULT 2
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neuroto
A;Reference number: S33411; MUID:93252228
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139

```
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin


  Query Match             15.3%;  Score 22;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-15;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      54 DNILFKIVGCNDTRYVGIRYFK 75
           ||||||||||||||||||||||
Db    1036 DNILFKIVGCNDTRYVGIRYFK 1057


RESULT   3
JH0256
botulinum neurotoxin type E precursor - Clostridium butyricum
C;Species: Clostridium butyricum
C;Date: 30-Jun-1992 #sequence_revision 15-May-1998 #text_change 16-Jul-1999
C;Accession: JH0256; S16145
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0256
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-27,'E',29-1251 <POU>
A;Cross-references: EMBL:X62088; NID:g40379
A;Experimental source: strains ATCC 43181 and ATCC 43755
R;Fujii, N.; Kimura, K.; Yashiki, T.; Indoh, T.; Murakami, T.; Tsuzuki, K.; Yokosawa, N.
J. Gen. Microbiol. 137, 519-525, 1991
A;Title: Cloning of a DNA fragment encoding the 5'-terminus of the botulinum type E toxi
A;Reference number: S16145; MUID:91237316
A;Accession: S16145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-229,'M',231-252 <FUJ>
A;Cross-references: EMBL:X53180; NID:g40407; PIDN:CAA37321.1; PID:g40408
A;Experimental source: strain BL6340
C;Comment: The clostridial neurotoxins are toxins that inhibit neurotransmitter release
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;2-422/Product: botulinum neurotoxin type E light chain #status predicted <LIG>
F;423-1251/Product: botulinum neurotoxin type E heavy chain #status predicted <HEA>
F;412-426/Disulfide bonds: #status predicted


  Query Match             10.4%;  Score 15;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           |||||||||||||||
Db     983 ISDYINKWIFVTITN 997


RESULT   4
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum t
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',96
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with oth
A;Reference number: A35294; MUID:90264400
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the
A;Reference number: A60027; MUID:90344918
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsi
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted


  Query Match             10.4%;  Score 15;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           |||||||||||||||
Db     983 ISDYINKWIFVTITN 997


RESULT   5
BTCLAB
bontoxilysin (EC 3.4.24.69) A precursor - Clostridium botulinum
N;Alternate names: botulinum neurotoxin type A
C;Species: Clostridium botulinum
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C;Accession: A35294; S09492; S68220; A33401; A53884; A60025; A27000
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with oth
A;Reference number: A35294; MUID:90264400
A;Accession: A35294
A;Molecule type: DNA
A;Residues: 1-1296 <BIN>
A;Cross-references: GB:M30196; NID:g144864; PIDN:AAA23262.1; PID:g144865
A;Experimental source: strain 62A, subtype A
R;Thompson, D.E.; Brehm, J.K.; Oultram, J.D.; Swinfield, T.J.; Shone, C.C.; Atkinson,
Eur. J. Biochem. 189, 73-81, 1990
A;Title: The complete amino acid sequence of the Clostridium botulinum type A neuroto
```

A;Reference number: S09492; MUID:90235864
A;Accession: S09492
A;Molecule type: DNA
A;Residues: 1,'Q',3-26,'V',28-1296 <THO>
A;Cross-references: EMBL:X52066; NID:g40381; PIDN:CAA36289.1; PID:g40382
A;Experimental source: NCTC 2916
R;Fujita, R.; Fujinaga, Y.; Inoue, K.; Nakajima, H.; Kumon, H.; Oguma, K.
FEBS Lett. 376, 41-44, 1995
A;Title: Molecular characterization of two forms of nontoxic-nonhemagglutinin components
A;Reference number: S67988; MUID:96096783
A;Accession: S68220
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-12 <FUJ>
A;Cross-references: EMBL:D67030; DDBJ:D50421; NID:g2160224
R;Betley, M.J.; Somers, E.; DasGupta, B.R.
Biochem. Biophys. Res. Commun. 162, 1388-1395, 1989
A;Title: Characterization of botulinum type A neurotoxin gene: delineation of the N-term
A;Reference number: A33401; MUID:89350959
A;Accession: A33401
A;Molecule type: DNA
A;Residues: 1-35 <BET>
A;Cross-references: GB:M27892; NID:g144880; PIDN:AAA23269.1; PID:g551776
R;Gimenez, J.A.; DasGupta, B.R.
J. Protein Chem. 12, 351-363, 1993
A;Title: Botulinum type A neurotoxin digested with pepsin yields 132, 97, 72, 45, 42, ar
A;Reference number: A53884; MUID:94000342
A;Accession: A53884
A;Status: preliminary
A;Molecule type: protein
A;Residues: 867-880;1148-1217,'Y',1219 <GIM>
A;Experimental source: strain Hall
A;Note: sequence extracted from NCBI backbone (NCBIP:139159); sequence modified after ex
R;DasGupta, B.R.; Dekleva, M.L.
Biochimie 72, 661-664, 1990
A;Title: Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and arou
A;Reference number: A60025; MUID:91120847
A;Accession: A60025
A;Molecule type: protein
A;Residues: 2-6;445-453,'X',455-457 <DAS1>
R;DasGupta, B.R.; Foley, J.; Niece, R.
Biochemistry 26, 4162, 1987
A;Title: Partial sequence of the light chain of botulinum neurotoxin type A.
A;Reference number: A27000
A;Accession: A27000
A;Molecule type: protein
A;Residues: 2-47 <DAS2>
R;Binz, T.; Blasi, J.; Yamasaki, S.; Baumeister, A.; Link, E.; Suedhof, T.C.; Jahn, R.;
J. Biol. Chem. 269, 1617-1620, 1994
A;Title: Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.
A;Reference number: A49708; MUID:94124495
A;Contents: annotation
C;Comment: Botulinum neurotoxins inhibit neurotransmitter release from cholinergic synap
C;Genetics:
A;Gene: atx; botA
C;Function:
A;Description: catalyzes hydrolysis of an Asn-Arg peptide bond in synaptosomal-associate
C;Superfamily: tetanus toxin
C;Keywords: disulfide bond; hydrolase; metalloproteinase; neurotoxin; transmembrane prot
F;2-444/Product: bontoxilysin A light chain #status experimental <LGHT>
F;445-1296/Product: bontoxilysin A heavy chain #status experimental <HVY>
F;223,227/Binding site: zinc (His) #status predicted
F;224/Active site: Glu #status predicted

Query Match 7.6%; Score 11; DB 1; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 WIFVTITNNRL 27
||||||||||||
Db 1014 WIFVTITNNRL 1024

RESULT 6
I40645
botulinum neurotoxin type A - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40645
R;Willems, A.; East, A.K.; Lawson, P.A.; Collins, M.D.
Res. Microbiol. 144, 547-556, 1993
A;Title: Sequence of the gene coding for the neurotoxin of Clostridium botulinum type
A;Reference number: I40645; MUID:94143603
A;Accession: I40645
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1296 <RES>
A;Cross-references: EMBL:X73423; NID:g507070; PIDN:CAA51824.1; PID:g507071
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 7.6%; Score 11; DB 2; Length 1296;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 WIFVTITNNRL 27
||||||||||||
Db 1014 WIFVTITNNRL 1024

RESULT 7
S48110
neurotoxin type F - Clostridium botulinum (fragment)
C;Species: Clostridium botulinum
C;Date: 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C;Accession: S48110
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48110
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 1-366 <CAM>
A;Cross-references: EMBL:X70821; NID:g407792; PIDN:CAA50152.1; PID:g407793
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 6.2%; Score 9; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VFNYTQMIS 9
|||||||||
Db 358 VFNYTQMIS 366

RESULT 8
B87350
hypothetical protein CC0813 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: B87350
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87350
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-540 <STO>

```
A;Cross-references: GB:AE005673; NID:g13422062; PIDN:AAK22798.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0813


  Query Match             5.6%;  Score 8;  DB 2;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      94 PDPSILKD 101
           ||||||||
Db      71 PDPSILKD 78


RESULT   9
S39791
neurotoxin - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 07-Oct-1994 #sequence_revision 01-Dec-1995 #text_change 16-Jul-1999
C;Accession: S39791
R;Campbell, K.; Collins, M.D.; East, A.K.
Biochim. Biophys. Acta 1216, 487-491, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium botulinum (Clostridium a
A;Reference number: S39791; MUID:94092745
A;Accession: S39791
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1297 <CAM>
A;Cross-references: EMBL:X74162; NID:g441275; PIDN:CAA52275.1; PID:g441276
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin


  Query Match             5.6%;  Score 8;  DB 2;  Length 1297;
  Best Local Similarity   100.0%;  Pred. No. 6.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKW 17
           ||||||||
Db    1002 ISDYINKW 1009


RESULT  10
S20990
myosin regulatory light chain 2 - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 16-Jul-1999
C;Accession: S20990
R;Alt, F.W.
submitted to the EMBL Data Library, May 1992
A;Reference number: S20990
A;Accession: S20990
A;Molecule type: mRNA
A;Residues: 1-156 <ALT>
A;Cross-references: EMBL:X65979
C;Superfamily: calmodulin; calmodulin repeat homology
C;Keywords: calcium binding; EF hand
F;14-46/Domain: calmodulin repeat homology <EF1>
F;84-116/Domain: calmodulin repeat homology <EF2>


  Query Match             4.9%;  Score 7;  DB 2;  Length 156;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db      93 FKVFDTE 99


RESULT  11
A81295
probable membrane protein Cj1484c [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 28-Jul-2000
C;Accession: A81295
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: A81295
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-200 <PAR>
A;Cross-references: GB:AL139078; GB:AL111168; NID:g6968723; PIDN:CAB73906.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj1484c
C;Superfamily: Campylobacter jejuni probable membrane protein Cj1484c


  Query Match             4.9%;  Score 7;  DB 2;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 KTEIETL 89
           |||||||
Db      66 KTEIETL 72


RESULT  12
S20992
myosin regulatory light chain 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 13-Aug-1999
C;Accession: S20992
R;Alt, F.W.
submitted to the EMBL Data Library, May 1992
A;Reference number: S20990
A;Accession: S20992
A;Molecule type: DNA
A;Residues: 1-202 <ALT>
A;Cross-references: EMBL:X65981; NID:g53747; PIDN:CAA46796.1; PID:g53748
C;Genetics:
A;Introns: 128/1
C;Superfamily: calmodulin; calmodulin repeat homology
C;Keywords: calcium binding; EF hand
F;60-92/Domain: calmodulin repeat homology <EF1>
F;130-162/Domain: calmodulin repeat homology <EF2>


  Query Match             4.9%;  Score 7;  DB 2;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db     139 FKVFDTE 145


RESULT  13
S22715
myosin regulatory light chain 2 - human
N;Alternate names: pre-lymphocyte-specific regulatory light chain PLRLC
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 20-Feb-1995 #text_change 16-Jul-1999
C;Accession: S22715
R;Oltz, E.M.; Yancopoulos, G.D.; Morrow, M.A.; Rolink, A.; Lee, G.; Wong, F.; Kaplan,
EMBO J. 11, 2759-2767, 1992
A;Title: A novel regulatory myosin light chain gene distinguishes pre-B cell subsets
A;Reference number: S22715; MUID:92331628
A;Accession: S22715
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-204 <OLT>
```

C;Superfamily: calmodulin; calmodulin repeat homology
C;Keywords: calcium binding; EF hand
F;62-94/Domain: calmodulin repeat homology <EF1>
F;132-164/Domain: calmodulin repeat homology <EF2>

Query Match 4.9%; Score 7; DB 2; Length 204;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 FKVFDTE 80
|||||||
Db 141 FKVFDTE 147

RESULT 14
AI2668
hypothetical protein Atu0751 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C;Accession: AI2668
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AI2668
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-226 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL41767.1; PID:g17739119; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu0751
A;Map position: circular chromosome

Query Match 4.9%; Score 7; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LGKTEIE 87
|||||||
Db 105 LGKTEIE 111

RESULT 15
S28494
dtxA protein - Clostridium difficile
N;Alternate names: tcdC
C;Species: Clostridium difficile
C;Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 15-Oct-1999
C;Accession: JC5344; S28494
R;Braun, V.; Hundsberger, T.; Leukel, P.; Sauerborn, M.; von Eichel-Streiber, C.
Gene 181, 29-38, 1996
A;Title: Definition of the single integration site of the pathogenicity locus in Clostri
A;Reference number: JC5340; MUID:97128764
A;Accession: JC5344
A;Molecule type: DNA
A;Residues: 1-231 <BRA>
A;Cross-references: EMBL:X92982; NID:g1770128; PIDN:CAA63565.1; PID:e212238; PID:g177013
A;Experimental source: strain VPI10463
C;Genetics:
A;Gene: dtxA; tcdC

Query Match 4.9%; Score 7; DB 2; Length 231;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 KTEIETL 89
|||||||
Db 70 KTEIETL 76

Search completed: August 15, 2002, 11:14:07
Job time: 259 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:38 ; Search time 24.69 Seconds
(without alignments)
225.825 Million cell updates/sec

Title: US-08-981-087A-3
Perfect score: 144
Sequence: 1 VFNYTQMISISDYINKWIFV.....ITQNSNFLNINQQRGVYQKP 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	27	18.8	1274	1	BXF_CLOBO	P30996 clostridium
2	15	10.4	1250	1	BXE_CLOBO	Q00496 clostridium
3	15	10.4	1250	1	BXE_CLOBU	P30995 clostridium
4	11	7.6	1295	1	BXA1_CLOBO	P10845 clostridium
5	11	7.6	1295	1	BXA2_CLOBO	Q45894 clostridium
6	8	5.6	1296	1	BXG_CLOBO	Q60393 clostridium
7	7	4.9	199	1	TDX2_THEAC	Q9hjl3 thermoplasm
8	7	4.9	267	1	EUTT_ECOLI	P76554 escherichia
9	7	4.9	267	1	EUTT_SALTY	Q9zfv4 salmonella
10	7	4.9	358	1	VAL1_CLVK	P14982 cassava lat
11	7	4.9	358	1	VAL1_CLVN	P14972 cassava lat
12	7	4.9	432	1	YZ21_METJA	Q60281 methanococc
13	7	4.9	508	1	YM05_ARCFU	O28078 archaeoglob
14	7	4.9	608	1	EDD_HELPJ	Q9zkb3 helicobacte
15	7	4.9	608	1	EDD_HELPY	P56111 helicobacte
16	7	4.9	665	1	ENV_MLVMO	P03385 moloney mur
17	7	4.9	814	1	PI3K_ARATH	P42339 arabidopsis
18	7	4.9	942	1	AMPN_MANSE	P91885 manduca sex
19	7	4.9	944	1	Y166_UREPA	Q9pqx7 ureaplasma
20	7	4.9	1035	1	ITA9_HUMAN	Q13797 homo sapien
21	7	4.9	1276	1	BXD_CLOBO	P19321 clostridium
22	7	4.9	1290	1	BXB_CLOBO	P10844 clostridium
23	7	4.9	1655	1	OMPB_RICCN	Q9kka3 r outer mem
24	7	4.9	1656	1	OMPB_RICJA	O06653 r outer mem
25	6	4.2	48	1	ATP8_ASPAM	P00858 aspergillus
26	6	4.2	48	1	ATP8_EMENI	P00857 emericella
27	6	4.2	49	1	Y195_BPT7	P03804 bacteriopha
28	6	4.2	63	1	RL29_BUCAK	P46174 buchnera ap
29	6	4.2	63	1	RL29_ECOLI	P02429 escherichia
30	6	4.2	74	1	YVFF_VACCC	P20564 vaccinia vi
31	6	4.2	75	1	YVAF_VACCC	P20515 vaccinia vi
32	6	4.2	78	1	HSTN_VIBCH	P04429 vibrio chol
33	6	4.2	78	1	HSTO_VIBCH	Q07425 vibrio chol
34	6	4.2	114	1	Y211_BUCAI	P57307 buchnera ap
35	6	4.2	128	1	WNT8_EVATR	P28093 evasterias
36	6	4.2	137	1	IPPD_PIG	Q29277 sus scrofa
37	6	4.2	148	1	AZUR_METFL	Q50400 methylobaci
38	6	4.2	151	1	YG98_CLOAB	P33661 clostridium
39	6	4.2	163	1	ATPX_OCHNE	Q40608 ochrosphaer
40	6	4.2	167	1	YLXL_BACSU	P40405 bacillus su
41	6	4.2	181	1	GLG2_HORVU	P55239 hordeum vul
42	6	4.2	189	1	YIWC_CAEEL	O01901 caenorhabdi
43	6	4.2	191	1	RERA_ARATH	O48670 arabidopsis
44	6	4.2	195	1	RERB_ARATH	O48671 arabidopsis
45	6	4.2	200	1	HBPM_RHIAP	O77422 rhipicephal

ALIGNMENTS

RESULT 1
BXF_CLOBO
ID BXF_CLOBO STANDARD; PRT; 1274 AA.
AC P30996;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE (Bontoxilysin F).
GN BOTF.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23387;
RX MEDLINE=93012902; PubMed=1398040;
RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA Roberts T.A., Thompson D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum.";
RL FEMS Microbiol. Lett. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=HOBBS FT10;
RX MEDLINE=94297488; PubMed=7764998;
RA East A.K., Collins M.D.;
RT "Conserved structure of genes encoding components of botulinum
RT neurotoxin complex M and the sequence of the gene coding for the
RT nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL Curr. Microbiol. 29:69-77(1994).
RN [3]
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC BOND OF SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No

```
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;


  Query Match             18.8%;  Score 27;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-20;
  Matches   27;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 YINKWIFVTITNNRLGNSRIYINGNLI 39
          ||||||||||||||||||||||||||
Db   1006 YINKWIFVTITNNRLGNSRIYINGNLI 1032


RESULT   2
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BELUGA;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; A60027; A60027.
```

```
DR   PIR; B35294; B35294.
DR   PIR; JH0257; JH0257.
DR   PIR; S08575; S08575.
DR   PIR; S18111; S18111.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;
```

```
  Query Match             10.4%;  Score 15;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           ||||||||||||||||
Db     982 ISDYINKWIFVTITN 996
```

```
RESULT   3
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, AND ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   PIR; S16145; S16145.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;
```

```
  Query Match             10.4%;  Score 15;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           ||||||||||||||||
Db     982 ISDYINKWIFVTITN 996
```

```
RESULT   4
BXA1_CLOBO
ID   BXA1_CLOBO     STANDARD;      PRT;  1295 AA.
AC   P10845; P18639; P01561;
DT   01-JUL-1989 (Rel. 11, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
```

```
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 2916;
RX   MEDLINE=90235864; PubMed=2185020;
RA   Thompson D.E., Brehm J.K., Oultram J.D., Swinfield T.-J.,
RA   Shone C.C., Atkinson T., Melling J., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type A
RT   neurotoxin, deduced by nucleotide sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 189:73-81(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz B., Kuarzono H., Wille M., Frevent J., Wernars K., Niemann H.;
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [3]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [4]
RP   SEQUENCE OF 1-34 FROM N.A.
RC   STRAIN=HALL;
RX   MEDLINE=89350959; PubMed=2669749;
RA   Betley M.J., Somers E., Dasgupta B.R.;
RT   "Characterization of botulinum type A neurotoxin gene: delineation of
RT   the N-terminal encoding region.";
RL   Biochem. Biophys. Res. Commun. 162:1388-1395(1989).
RN   [5]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
RN   [6]
RP   SEQUENCE OF 1-16.
RX   MEDLINE=84178501; PubMed=6370252;
RA   Schmidt J.J., Sartymoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequence of the heavy and light chains of
RT   botulinum neurotoxin type A.";
RL   Biochem. Biophys. Res. Commun. 119:900-904(1984).
RN   [7]
RP   SEQUENCE OF 1-46.
RA   Dasgupta B.R., Foley J., Niece R.;
RT   "Partial sequence of the light chain of botulinum neurotoxin type A.";
RL   Biochemistry 26:4162-4162(1987).
RN   [8]
RP   SEQUENCE OF 1-5 AND 444-456.
RX   MEDLINE=91120847; PubMed=2126206;
RA   Dasgupta B.R., Dekleva M.L.;
RT   "Botulinum neurotoxin type A: sequence of amino acids at the
RT   N-terminus and around the nicking site.";
RL   Biochimie 72:661-664(1990).
RN   [9]
RP   SEQUENCE OF 448-464 AND 872-895.
RX   MEDLINE=89024662; PubMed=3178218;
RA   Sathymoorthy V., Dasgupta B.R., Foley J., Niece R.L.;
RT   "Botulinum neurotoxin type A: cleavage of the heavy chain into two
RT   halves and their partial sequences.";
RL   Arch. Biochem. Biophys. 266:142-151(1988).
RN   [10]
RP   SEQUENCE OF 448-482.
RX   MEDLINE=85285016; PubMed=3896784;
RA   Shone C.C., Hambleton P., Melling J.;
RT   "Inactivation of Clostridium botulinum type A neurotoxin by trypsin
RT   and purification of two tryptic fragments. Proteolytic action near
RT   the COOH-terminus of the heavy subunit destroys toxin-binding
RT   activity.";
RL   Eur. J. Biochem. 151:75-82(1985).
RN   [11]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [12]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
RN   [13]
RP   MUTAGENESIS OF GLU-261; PHE-265 AND TYR-365.
RX   PubMed=11700044;
RA   Rigoni M., Caccin P., Johnson E.A., Montecucco C., Rossetto O.;
RT   "Site-directed mutagenesis identifies active-site residues of the
RT   light chain of botulinum neurotoxin type a.";
RL   Biochem. Biophys. Res. Commun. 288:1231-1237(2001).
RN   [14]
RP   X-RAY CRYSTALLOGRAPHY (3.3 ANGSTROMS).
RX   MEDLINE=98455071; PubMed=9783750;
RA   Lacy D.B., Tepp W., Cohen A.C., Dasgupta B.R., Stevens R.C.;
RT   "Crystal structure of botulinum neurotoxin type A and implications
RT   for toxicity.";
RL   Nat. Struct. Biol. 5:898-902(1998).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- PHARMACEUTICAL: Available under the name BOTOX(R) (Allergan) for
CC       the treatment of strabismus and blepharospasm associated with
CC       dystonia and cervical dystonia. Also used for the treatment of
CC       hemifacial spasm and a number of other neurological disorders
CC       characterized by abnormal muscle contraction.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X52066; CAA36289.1; -.
DR   EMBL; M30196; AAA23262.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   EMBL; M27892; AAA23269.1; -.
DR   PIR; A35294; BTCLAB.
DR   PIR; S09492; S09492.
DR   PDB; 3BTA; 01-OCT-99.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc;
KW   Pharmaceutical; 3D-structure.
FT   INIT_MET      0      0
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC).
FT   ACT_SITE    223    223
FT   METAL       226    226       ZINC (CATALYTIC).
FT   METAL       261    261       ZINC (CATALYTIC).
FT   DISULFID    429    453       INTERCHAIN.
FT   DISULFID   1234   1279
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
FT   VARIANT      26     26       V -> A.
FT   MUTAGEN     261    261       E->A: DRASTIC DECREASE IN ENZYMATIC
FT                                ACTIVITY.
FT   MUTAGEN     265    265       F->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   MUTAGEN     365    365       Y->A: DECREASE IN ENZYMATIC ACTIVITY.
FT   CONFLICT      1      1       P -> Q (IN REF. 1).
FT   CONFLICT    479    479       E -> P (IN REF. 9).
FT   CONFLICT    875    875       T -> L (IN REF. 8).
FT   CONFLICT    891    891       S -> K (IN REF. 8).
SQ   SEQUENCE   1295 AA;  149322 MW;  858342F754862579 CRC64;


  Query Match             7.6%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 WIFVTITNNRL 27
           ||||||||||| 
Db    1013 WIFVTITNNRL 1023


RESULT   5
BXA2_CLOBO
ID   BXA2_CLOBO     STANDARD;      PRT;  1295 AA.
AC   Q45894; P77780;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type A precursor (EC 3.4.24.69) (BoNT/A)
DE   (Bontoxilysin A) (BOTOX) [Contains: Botulinum neurotoxin A, light-
DE   chain; Botulinum neurotoxin A, heavy-chain].
GN   BOTA OR BNA OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=Kyoto-F;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
CC   -!- FUNCTION: Inhibits acetylcholine release. The botulinum toxin
CC       binds with high affinity to peripheral neuronal presynaptic
CC       membrane, is then internalized by receptor-mediated endocytosis.
CC       The C-terminus of the heavy chain (H) is responsible for the
CC       adherence of the toxin to the cell surface while the N-terminus
CC       mediates transport of the light chain from the endocytic vesicle
CC       to the cytosol. After translocation, the light chain (L)
CC       hydrolyzes the 197-Gln-|-Arg-198 bond in SNAP-25, thereby blocking
CC       neurotransmitter release. Inhibition of acetylcholine release
CC       results in flaccid paralysis, with frequent heart or respiratory
CC       failure (By similarity).
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: Disulfide-linked heterodimer of a light chain (L) and a
CC       heavy chain (H) (By similarity).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: There are seven antigenically distinct forms of
CC       botulinum neurotoxin: Types A, B, C1, D, E, F, and G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    447       BOTULINUM NEUROTOXIN A, LIGHT-CHAIN.
FT   CHAIN       448   1295       BOTULINUM NEUROTOXIN A, HEAVY-CHAIN.
FT   METAL       222    222       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    223    223       BY SIMILARITY.
FT   METAL       226    226       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    453       INTERCHAIN (BY SIMILARITY).
FT   DISULFID   1234   1279       BY SIMILARITY.
FT   TRANSMEM    626    646       POTENTIAL.
FT   TRANSMEM    655    675       POTENTIAL.
SQ   SEQUENCE   1295 AA;  149279 MW;  5DA04A13D98D6372 CRC64;


  Query Match             7.6%;  Score 11;  DB 1;  Length 1295;
  Best Local Similarity   100.0%;  Pred. No. 0.0021;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 WIFVTITNNRL 27
           ||||||||||| 
Db    1013 WIFVTITNNRL 1023
```

```
RESULT   6
BXG_CLOBO
ID   BXG_CLOBO      STANDARD;      PRT;  1296 AA.
AC   Q60393;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type G precursor (EC 3.4.24.69) (BoNT/G)
DE   (Bontoxilysin G).
GN   BOTG.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=113 / 30;
RX   MEDLINE=94092745; PubMed=8268233;
RA   Campbell K., Collins M.D., East A.K.;
RT   "Nucleotide sequence of the gene coding for Clostridium botulinum
RT   (Clostridium argentinense) type G neurotoxin: genealogical comparison
RT   with other clostridial neurotoxins.";
RL   Biochim. Biophys. Acta 1216:487-491(1993).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74162; CAA52275.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   CHAIN         1    441       BOTULINUM NEUROTOXIN G, LIGHT-CHAIN.
FT   CHAIN       442   1296       BOTULINUM NEUROTOXIN G, HEAVY-CHAIN.
FT   METAL       229    229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    230    230       BY SIMILARITY.
FT   METAL       233    233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    435    449       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1296 AA;  149013 MW;  DC8E47E15F665C31 CRC64;


  Query Match             5.6%;  Score 8;  DB 1;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 2.8;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKW 17
           ||||||||
Db    1001 ISDYINKW 1008


RESULT   7
TDX2_THEAC
ID   TDX2_THEAC     STANDARD;      PRT;   199 AA.
AC   Q9HJL3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable peroxiredoxin 2.
GN   TA0954.
OS   Thermoplasma acidophilum.
OC   Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
OC   Thermoplasma.
OX   NCBI_TaxID=2303;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DSM 1728;
RX   MEDLINE=20479972; PubMed=11029001;
RA   Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA   Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT   "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT   acidophilum.";
RL   Nature 407:508-513(2000).
CC   -!- SIMILARITY: BELONGS TO THE AHPC/TSA FAMILY. TDXH SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL445066; CAC12083.1; -.
DR   InterPro; IPR000866; AhpC-TSA.
DR   Pfam; PF00578; AhpC-TSA; 1.
KW   Antioxidant; Complete proteome.
FT   ACT_SITE     40     40       BY SIMILARITY.
SQ   SEQUENCE   199 AA;  22581 MW;  4E266162F5D58162 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 199;
  Best Local Similarity   100.0%;  Pred. No. 5.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      37 NLIDEKS 43
           |||||||
Db     104 NLIDEKS 110


RESULT   8
EUTT_ECOLI
ID   EUTT_ECOLI     STANDARD;      PRT;   267 AA.
AC   P76554;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).
GN   EUTT OR B2459.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
```

```
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
CC   -!- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + cob(I)alamin + H(2)O = phosphate +
CC       diphosphate + adenosylcobalamin.
CC   -!- PATHWAY: ETHANOLAMINE UTILIZATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000332; AAC75512.1; -.
DR   EcoGene; EG14189; eutT.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   267 AA;  30171 MW;  E51EDAB528B4FA76 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 7.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     115 LLNLLRT 121
           |||||||
Db     199 LLNLLRT 205


RESULT   9
EUTT_SALTY
ID   EUTT_SALTY     STANDARD;      PRT;   267 AA.
AC   Q9ZFV4;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Ethanolamine utilization cobalamin adenosyltransferase (EC 2.5.1.17).
GN   EUTT OR STM2467.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=99395039; PubMed=10464203;
RA   Kofoid E.C., Rappleye C.A., Stojiljkovic I., Roth J.R.;
RT   "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium
RT   encodes five homologues of carboxysome shell proteins.";
RL   J. Bacteriol. 181:5317-5329(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- FUNCTION: CONVERTS CNB12 TO ADOB12 (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + cob(I)alamin + H(2)O = phosphate +
CC       diphosphate + adenosylcobalamin.
CC   -!- PATHWAY: ETHANOLAMINE UTILIZATION.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF093749; AAC78114.1; -.
DR   EMBL; AE008811; AAL21361.1; -.
DR   StyGene; SG10636; eutT.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   267 AA;  30238 MW;  9502A28FDB4DC9E4 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 267;
  Best Local Similarity   100.0%;  Pred. No. 7.4;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     115 LLNLLRT 121
           |||||||
Db     199 LLNLLRT 205


RESULT  10
VAL1_CLVK
ID   VAL1_CLVK      STANDARD;      PRT;   358 AA.
AC   P14982;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   AL1 protein (40.4 kDa protein).
GN   AC1.
OS   Cassava latent virus (strain West Kenyan 844).
OC   Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX   NCBI_TaxID=10818;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stanley J., Gay M.R.;
RT   "Nucleotide sequence of cassava latent virus DNA.";
RL   Nature 301:260-262(1983).
CC   -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02057; -; NOT_ANNOTATED_CDS.
DR   InterPro; IPR001191; Gemini_AL1.
DR   Pfam; PF00799; Gemini_AL1; 1.
DR   PRINTS; PR00227; GEMCOATAL1.
DR   ProDom; PD000736; Gemini_AL1; 1.
KW   ATP-binding.
FT   NP_BIND     220    227       ATP (POTENTIAL).
SQ   SEQUENCE   358 AA;  40346 MW;  ED173E753EE92D69 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 TITNNRL 27
           |||||||
Db      68 TITNNRL 74


RESULT  11
VAL1_CLVN
ID   VAL1_CLVN      STANDARD;      PRT;   358 AA.
AC   P14972;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
```

```
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   AL1 protein (40.4 kDa protein).
GN   AC1.
OS   Cassava latent virus (strain Nigerian).
OC   Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX   NCBI_TaxID=10819;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90174930; PubMed=2308831;
RA   Morris B., Coates L., Lowe S., Richardson K., Eddy P.;
RT   "Nucleotide sequence of the infectious cloned DNA components of
RT   African cassava mosaic virus (Nigerian strain).";
RL   Nucleic Acids Res. 18:197-198(1990).
CC   -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X17095; CAA34953.1; -.
DR   PIR; S07594; S07594.
DR   InterPro; IPR001191; Gemini_AL1.
DR   Pfam; PF00799; Gemini_AL1; 1.
DR   PRINTS; PR00227; GEMCOATAL1.
DR   ProDom; PD000736; Gemini_AL1; 1.
KW   ATP-binding.
FT   NP_BIND     220    227       ATP (POTENTIAL).
SQ   SEQUENCE   358 AA;  40435 MW;  1DB16BB0CB2D5E2C CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 358;
  Best Local Similarity   100.0%;  Pred. No. 9.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      21 TITNNRL 27
           |||||||
Db      68 TITNNRL 74


RESULT  12
YZ21_METJA
ID   YZ21_METJA     STANDARD;      PRT;   432 AA.
AC   Q60281;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJECL21.
GN   MJECL21.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC   Methanococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L77118; AAC37092.1; -.
DR   TIGR; MJECL21; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   432 AA;  51081 MW;  DBADF2C5C43A4F90 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     114 YLLNLLR 120
           |||||||
Db     211 YLLNLLR 217


RESULT  13
YM05_ARCFU
ID   YM05_ARCFU     STANDARD;      PRT;   508 AA.
AC   O28078;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein AF2205.
GN   AF2205.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000953; AAB89056.1; -.
DR   TIGR; AF2205; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM      7     29       POTENTIAL.
SQ   SEQUENCE   508 AA;  56562 MW;  85823142F601FC6D CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 508;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      22 ITNNRLG 28
           |||||||
Db     373 ITNNRLG 379


RESULT  14
EDD_HELPJ
ID   EDD_HELPJ      STANDARD;      PRT;   608 AA.
AC   Q9ZKB3;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphogluconate dehydratase (EC 4.2.1.12) (6-phosphogluconate
DE   dehydratase).
GN   EDD OR JHP1026.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- CATALYTIC ACTIVITY: 6-phospho-D-gluconate = 2-dehydro-3-deoxy-6-
CC       phospho-D-gluconate + H(2)O.
CC   -!- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001530; AAD06597.1; -.
DR   InterPro; IPR000581; ILVD_EDD.
DR   Pfam; PF00920; ILVD_EDD; 1.
DR   ProDom; PD002691; ILVD_EDD; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
DR   PROSITE; PS00887; ILVD_EDD_2; 1.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   608 AA;  66603 MW;  978A046F3AE15F98 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 LIDEKSI 44
           |||||||
Db     276 LIDEKSI 282


RESULT  15
EDD_HELPY
ID   EDD_HELPY      STANDARD;      PRT;   608 AA.
AC   P56111;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphogluconate dehydratase (EC 4.2.1.12) (6-phosphogluconate
DE   dehydratase).
GN   EDD OR HP1100.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- CATALYTIC ACTIVITY: 6-phospho-D-gluconate = 2-dehydro-3-deoxy-6-
CC       phospho-D-gluconate + H(2)O.
CC   -!- PATHWAY: KEY ENZYME IN THE ENTNER-DOUDOROFF PATHWAY.
CC   -!- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000616; AAD08143.1; -.
DR   TIGR; HP1100; -.
DR   InterPro; IPR000581; ILVD_EDD.
DR   Pfam; PF00920; ILVD_EDD; 1.
DR   ProDom; PD002691; ILVD_EDD; 1.
DR   PROSITE; PS00886; ILVD_EDD_1; 1.
DR   PROSITE; PS00887; ILVD_EDD_2; 1.
KW   Lyase; Complete proteome.
SQ   SEQUENCE   608 AA;  66655 MW;  47EF7E62E3371F59 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 608;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      38 LIDEKSI 44
           |||||||
Db     276 LIDEKSI 282


Search completed: August 15, 2002, 11:24:39
Job time: 686 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:27 ; Search time 96.53 Seconds
(without alignments)
164.545 Million cell updates/sec

Title: US-08-981-087A-4
Perfect score: 143
Sequence: 1 NIFSNTRLYTGVEVIIRKNG.....TSSNGCFWSFISKEHGWQEN 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	143	100.0	143	18	AAW09017	Immunogenic type F
2	143	100.0	431	18	AAW09014	Immunogenic type F
3	143	100.0	432	22	AAB04096	Botulism toxin hea
4	143	100.0	432	22	AAB04103	Botulism toxin hea
5	143	100.0	645	22	AAE07894	Modified clostridi
6	143	100.0	685	22	AAE07893	Modified clostridi
7	143	100.0	862	22	AAE07890	Modified clostridi
8	143	100.0	887	22	AAE07892	Modified clostridi
9	143	100.0	1032	22	AAE07901	C. botulinum C2 tr
10	143	100.0	1059	21	AAY93309	A manganese supero
11	143	100.0	1084	21	AAY93312	A manganese supero
12	143	100.0	1092	22	AAE07900	C. botulinum C2 tr
13	92	64.3	432	21	AAY77138	Synthetic botulinu
14	24	16.8	448	19	AAW68399	Clostridium botuli
15	8	5.6	419	22	AAB04095	Botulism toxin hea
16	8	5.6	449	21	AAY77137	Synthetic botulinu
17	8	5.6	449	22	AAB04094	Botulism toxin hea
18	8	5.6	452	19	AAW68396	Clostridium botuli
19	7	4.9	96	21	AAB69310	HIV-1 non-subtype
20	7	4.9	104	22	AAO06221	Human polypeptide
21	7	4.9	114	22	AAE10807	Xanthomonas campes
22	7	4.9	245	20	AAY20027	B. burgdorferi ant
23	7	4.9	265	20	AAY20026	B. burgdorferi ant
24	7	4.9	289	22	AAG82013	S. epidermidis ope
25	7	4.9	335	21	AAY95073	Candida albicans p
26	7	4.9	387	12	AAR12576	Ad41 long fibre pr
27	7	4.9	387	13	AAR23652	Tak short fibre pr
28	7	4.9	451	19	AAW68395	Clostridium botuli
29	7	4.9	518	22	ABB63292	Drosophila melanog
30	7	4.9	554	19	AAW46873	Bacillus thuringie
31	7	4.9	554	20	AAY24976	Bacillus thuringie
32	7	4.9	692	20	AAW93890	Human HG38 protein
33	7	4.9	693	20	AAY42169	Human LGR5 protein
34	7	4.9	797	22	AAU33386	Enterococcus faeca
35	7	4.9	798	22	AAU35171	Enterococcus faeca
36	7	4.9	819	22	ABB67658	Drosophila melanog
37	7	4.9	907	20	AAW93889	Human HG38 protein
38	7	4.9	907	21	AAY90682	Human G protein-co
39	7	4.9	907	21	AAY90687	Human mutant G pro
40	7	4.9	1004	22	ABB64051	Drosophila melanog
41	7	4.9	1049	22	ABB65142	Drosophila melanog
42	7	4.9	1171	22	ABB64387	Drosophila melanog
43	7	4.9	1265	22	ABG00708	Novel human diagno
44	7	4.9	1284	22	ABB71869	Drosophila melanog
45	7	4.9	2280	22	ABB61650	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW09017
ID AAW09017 standard; Protein; 143 AA.
XX
AC AAW09017;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin polypeptide (aa1136-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 19; 37pp; English.
XX
CC Novel polypeptides (AAW09014-17) respectively comprise amino acids

CC 848-1278, 848-991, 992-1135 and 1136-1278 in the heavy chain of a
CC type F botulinum neurotoxin (BoNT/F). They lack the L chain and
CC HN epitopes necessary for metalloprotease activity and toxin
CC internalisation. They are free of botulinum toxin activity but can
CC induce protective immunity to a type F botulinum toxin, making them
CC useful for vaccine prodn. Recombinant polypeptides can be
CC produced in transformed host cells, esp. as fusion proteins, e.g.
CC with maltose binding protein to facilitate purification.
XX
SQ Sequence 143 AA;

Query Match 100.0%; Score 143; DB 18; Length 143;
Best Local Similarity 100.0%; Pred. No. 5.7e-140;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
Db 1 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 60

Qy 61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
Db 61 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 120

Qy 121 RKNTSSNGCFWSFISKEHGWQEN 143
Db 121 rkntssngcfwsfiskehgwqen 143

RESULT 2
AAW09014
ID AAW09014 standard; Protein; 431 AA.
XX
AC AAW09014;
XX
DT 31-MAR-1997 (first entry)
XX
DE Immunogenic type F botulinum toxin heavy chain (aa848-1278).
XX
KW Botulinum toxin; neurotoxin; BoBT/F; immunogen; vaccine; botulism.
XX
OS Clostridium botulinum type F strain Langeland.
XX
PN WO9641881-A1.
XX
PD 27-DEC-1996.
XX
PF 12-JUN-1996; 96WO-GB01409.
XX
PR 12-JUN-1995; 95GB-0011909.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Elmore MJ, Mauchline ML, Minton NP, Pasechnik VA;
XX
DR WPI; 1997-065467/06.
DR N-PSDB; AAT48100.
XX
PT Immunogenic type F botulinum toxin polypeptide(s) - allows
PT recombinant vaccine prodn.
XX
PS Claim 5; Page 16-17; 37pp; English.
XX
CC A polypeptide (AAW09014) comprises the heavy chain (amino acids
CC 848-1278) of a type F botulinum neurotoxin (BoNT/F), and can be
CC produced using a synthetic gene (AAT48101) based on the natural
CC gene sequence (AAT48100) for the heavy chain. The polypeptides and
CC its fragments (see also AAW09015-17) lack the light chain and HN
CC epitopes necessary for metalloprotease activity and toxin
CC internalisation. They are free of botulinum toxin activity but can
CC induce protective immunity to a type F botulinum toxin, making them
CC useful for vaccine prodn. Recombinant polypeptides can be
CC produced in transformed host cells, esp. as fusion proteins, e.g.
CC with maltose binding protein to facilitate purification.
XX
SQ Sequence 431 AA;

Query Match 100.0%; Score 143; DB 18; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
Db 289 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 348

Qy 61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
Db 349 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 408

Qy 121 RKNTSSNGCFWSFISKEHGWQEN 143
Db 409 rkntssngcfwsfiskehgwqen 431

RESULT 3
AAB04096
ID AAB04096 standard; Protein; 432 AA.
XX
AC AAB04096;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54490.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Claim 3; Fig 9b; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce

```
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match             100.0%;  Score 143;  DB 22;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-139;
  Matches  143;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 349

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 409

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    410 rkntssngcfwsfiskehgwqen 432


RESULT   4
AAB04103
ID   AAB04103 standard; Protein; 432 AA.
XX
AC   AAB04103;
XX
DT   11-APR-2001  (first entry)
XX
DE   Botulism toxin heavy chain C-terminal sequence (serotype F).
XX
KW   Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW   recombinant vector; antigen; immune response; vaccine; bacterium;
KW   infection.
XX
OS   Synthetic.
OS   Clostridium botulinum.
XX
PN   WO200067700-A2.
XX
PD   16-NOV-2000.
XX
PF   12-MAY-2000; 2000WO-US12890.
XX
PR   12-MAY-1999;   99US-0133865.
PR   12-MAY-1999;   99US-0133866.
PR   12-MAY-1999;   99US-0133867.
PR   12-MAY-1999;   99US-0133868.
PR   12-MAY-1999;   99US-0133869.
PR   12-MAY-1999;   99US-0133873.
PR   29-JUL-1999;   99US-0146192.
XX
PA   (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI   Smith LA,  Byrne MP,  Middlebrook JL,  Lapenotiere H;
XX
DR   WPI; 2001-016048/02.
DR   N-PSDB; AAA54499.
XX
PT   New nucleic acids encoding the carboxy- or amino-terminal portions of
PT   the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT   vaccine against botulism
XX
PS   Disclosure; Fig 18b; 73pp; English.
XX
CC   Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC   chain and then posttranslationally nicked, forming a dichain
CC   consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC   remain linked by a disulfide bond. Nucleic acids encoding the
CC   carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC   chain  of botulinum neurotoxin (BoNT) can be used in recombinant
CC   expression vectors and expressed in transformed cells to produce
CC   peptide antigens useful for eliciting an immune response to give
CC   protective immunity against botulinum neurotoxin, which causes
CC   botulism. The nucleic acids are expressible in a recombinant
CC   organisms such as Escherichia coli or Pichia pastoris. The use
CC   of recombinant nucleic acids are advantageous since it eliminates
CC   the need to culture large quantities of hazardous toxin-producing
CC   bacterium. Production yield from the genetically engineered product
CC   is also high and cost of production is lower. The nucleic acids can
CC   be derived from Clostridium botulinum serotypes A-G.
XX
SQ   Sequence   432 AA;


  Query Match             100.0%;  Score 143;  DB 22;  Length 432;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-139;
  Matches  143;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    290 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 349

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    350 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 409

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    410 rkntssngcfwsfiskehgwqen 432


RESULT   5
AAE07894
ID   AAE07894 standard; Protein; 645 AA.
XX
AC   AAE07894;
XX
DT   01-NOV-2001  (first entry)
XX
DE   Modified clostridial heavy chain fragment #1.
XX
KW   Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW   tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW   diphtheria neurotoxin; botulinum neurotoxin type F; BoNT/F.
XX
OS   Chimeric - Corynebacterium diphtheriae.
OS   Chimeric - Clostridium botulinum.
XX
PN   WO200158936-A2.
XX
PD   16-AUG-2001.
XX
PF   04-DEC-2000; 2000WO-GB04644.
XX
PR   02-DEC-1999;   99GB-0028530.
PR   07-APR-2000; 2000GB-0008658.
XX
PA   (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI   Shone CC,  Sutton JM,  Silman N;
XX
DR   WPI; 2001-514643/56.
XX
PT   New non toxic polypeptide for delivery of a therapeutic agent for the
PT   treatment of a CNS disorder comprising a binding domain that
PT   translocates the therapeutic agent into the neuronal cells -
```

XX
PS Example 2; Page 44; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain fragment. This sequence is
CC constructed by fusing the binding domain of botulinum neurotoxin type F
CC (BoNT/F) with translocation domain of diphtheria neurotoxin.
XX
SQ Sequence 645 AA;

Query Match 100.0%; Score 143; DB 22; Length 645;
Best Local Similarity 100.0%; Pred. No. 2.2e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    503 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 562

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    563 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 622

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    623 rkntssngcfwsfiskehgwqen 645
```

RESULT 6
AAE07893
ID AAE07893 standard; Protein; 685 AA.
XX
AC AAE07893;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #5.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Influenza virus.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 43; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate. This
CC conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation peptide from influenza virus and a neuronal cell-specific
CC binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 685 AA;

Query Match 100.0%; Score 143; DB 22; Length 685;
Best Local Similarity 100.0%; Pred. No. 2.3e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    543 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 602

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 662

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    663 rkntssngcfwsfiskehgwqen 685
```

RESULT 7
AAE07890
ID AAE07890 standard; Protein; 862 AA.
XX
AC AAE07890;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.

PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 40; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate.
CC This conjugate comprises bacterial Mn-superoxide dismutase (MnSOD), from
CC Bacillus stearothermophilus, linker that can be cleaved by factor Xa,
CC translocation domain from diphtheria neurotoxin and a neuronal cell-
CC specific binding domain from botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 862 AA;

Query Match 100.0%; Score 143; DB 22; Length 862;
Best Local Similarity 100.0%; Pred. No. 2.9e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    720 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 779

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    780 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 839

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    840 rkntssngcfwsfiskehgwqen 862
```

RESULT 8
AAE07892
ID AAE07892 standard; Protein; 887 AA.
XX
AC AAE07892;
XX
DT 01-NOV-2001 (first entry)
XX
DE Modified clostridial heavy chain-superoxide dismutase conjugate #4.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy; chimeric;
KW superoxide dismutase; SOD; diphtheria neurotoxin; human;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Bacillus stearothermophilus.
OS Chimeric - Corynebacterium diphtheriae.
OS Chimeric - Clostridium botulinum.
OS Chimeric - Synthetic.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 9; Page 42; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is modified clostridial heavy chain-superoxide dismutase conjugate.
CC This conjugate comprises a mitochondrial leader sequence from human
CC Mn-superoxide dismutase (MnSOD), MnSOD from Bacillus stearothermophilus,
CC linker that can be cleaved by thrombin, translocation domain from
CC diphtheria neurotoxin and a neuronal cell-specific binding domain from
CC botulinum neurotoxin type F (BoNT/F).
XX
SQ Sequence 887 AA;

Query Match 100.0%; Score 143; DB 22; Length 887;
Best Local Similarity 100.0%; Pred. No. 3e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    745 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 804

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    805 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 864

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    865 rkntssngcfwsfiskehgwqen 887
```

RESULT 9
AAE07901
ID AAE07901 standard; Protein; 1032 AA.
XX
AC AAE07901;
XX
DT 01-NOV-2001 (first entry)
XX
DE C. botulinum C2 translocation domain with BoNT/F-binding domain #2.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;

KW tumour; infection; neurodegenerative disease; gene therapy;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 48; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC the invention.
XX
SQ Sequence 1032 AA;

Query Match 100.0%; Score 143; DB 22; Length 1032;
Best Local Similarity 100.0%; Pred. No. 3.4e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     890 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 949

Qy      61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggniglgfhsnnlvasswyynni 1009

Qy     121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    1010 rkntssngcfwsfiskehgwqen 1032
```

RESULT 10
AAY93309
ID AAY93309 standard; protein; 1059 AA.
XX
AC AAY93309;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX

KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB03699.
XX
PR 05-NOV-1998; 98GB-0024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS Disclosure; Page 48-51; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a manganese superoxide dismutase (Mn-SOD) polypeptide, a linker that
CC can be cleaved by thrombin, and a heavy chain derived from botulinum
CC neurotoxin serotype F. The specification describes a composition for
CC delivery of SOD to neuronal cells. The composition comprises SOD linked,
CC by a cleavable linker, to a neuronal cell targeting component (NCTC).
CC This component has a domain that binds to a neuronal cell and a
CC domain that translocates the SOD of the composition into the neuronal
CC cell. After translocation, the linker is cleaved to release the SOD.
CC The composition is useful for treating neuronal diseases caused or
CC augmented by oxidative stress, such as ischemic stroke, trauma,
CC Parkinson's disease, Huntington's disease and motor neurone diseases.
XX
SQ Sequence 1059 AA;

Query Match 100.0%; Score 143; DB 21; Length 1059;
Best Local Similarity 100.0%; Pred. No. 3.5e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     917 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 976

Qy      61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     977 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggniglgfhsnnlvasswyynni 1036

Qy     121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db    1037 rkntssngcfwsfiskehgwqen 1059
```

RESULT 11
AAY93312
ID AAY93312 standard; protein; 1084 AA.
XX
AC AAY93312;
XX
DT 04-SEP-2000 (first entry)
XX
DE A manganese superoxide dismutase (Mn-SOD) construct.
XX

KW Manganese superoxide dismutase; Mn-SOD; SOD; neuronal cell;
KW neuronal cell targeting component; NCTC; neuronal disease;
KW oxidative stress; ischemic stroke; trauma; Parkinson's disease;
KW Huntington's disease; motor neurone disease;
KW botulinum neurotoxin serotype F.
XX
OS Synthetic.
OS Homo sapiens.
OS Bacillus stearothermophilus.
OS Clostridium botulinum.
XX
PN WO200028041-A1.
XX
PD 18-MAY-2000.
XX
PF 05-NOV-1999; 99WO-GB03699.
XX
PR 05-NOV-1998; 98GB-0024282.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Hallis B, Silman N;
XX
DR WPI; 2000-376553/32.
XX
PT Novel composition, comprising superoxide dismutase linked by a
PT cleavable linker to a neuronal cell targeting component useful for
PT delivering superoxide dismutase to neuronal cells to treat ischemia -
XX
PS Disclosure; Page 57-60; 65pp; English.
XX
CC The present sequence represents a construct of the invention, comprising
CC a mitochondrial leader sequence from human manganese superoxide
CC dismutase (Mn-SOD), a Bacillus stearothermophilus Mn-SOD, a linker
CC that can be cleaved by thrombin, and a heavy chain derived from
CC botulinum neurotoxin serotype F. The specification describes a
CC composition for delivery of SOD to neuronal cells. The composition
CC comprises SOD linked, by a cleavable linker, to a neuronal cell
CC targeting component (NCTC). This component has a domain that binds
CC to a neuronal cell and a domain that translocates the SOD of the
CC composition into the neuronal cell. After translocation, the linker
CC is cleaved to release the SOD. The composition is useful for treating
CC neuronal diseases caused or augmented by oxidative stress, such as
CC ischemic stroke, trauma, Parkinson's disease, Huntington's disease and
CC motor neurone diseases.
XX
SQ Sequence 1084 AA;

Query Match 100.0%; Score 143; DB 21; Length 1084;
Best Local Similarity 100.0%; Pred. No. 3.6e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 942 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 1001

Qy 61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1002 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 1061

Qy 121 RKNTSSNGCFWSFISKEHGWQEN 143
|||||||||||||||||||||||
Db 1062 rkntssngcfwsfiskehgwqen 1084

RESULT 12
AAE07900
ID AAE07900 standard; Protein; 1092 AA.
XX
AC AAE07900;
XX
DT 01-NOV-2001 (first entry)
XX
DE C. botulinum C2 translocation domain with BoNT/F-binding domain #1.
XX
KW Neuronal cell; binding domain; translocation domain; stroke; epilepsy;
KW tumour; infection; neurodegenerative disease; gene therapy;
KW botulinum neurotoxin type F; BoNT/F.
XX
OS Clostridium botulinum.
XX
PN WO200158936-A2.
XX
PD 16-AUG-2001.
XX
PF 04-DEC-2000; 2000WO-GB04644.
XX
PR 02-DEC-1999; 99GB-0028530.
PR 07-APR-2000; 2000GB-0008658.
XX
PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX
PI Shone CC, Sutton JM, Silman N;
XX
DR WPI; 2001-514643/56.
XX
PT New non toxic polypeptide for delivery of a therapeutic agent for the
PT treatment of a CNS disorder comprising a binding domain that
PT translocates the therapeutic agent into the neuronal cells -
XX
PS Example 2; Page 47; 50pp; English.
XX
CC The invention relates to a non toxic polypeptide, for delivery of a
CC therapeutic agent to a neuronal cell, which comprises a binding domain
CC (carboxy terminal half of heavy chain (HC) of a neurotoxin, designated
CC as Hc) that binds to the neuronal cell and a translocation domain (amino
CC terminal half of HC, designated as HN), that translocates the therapeutic
CC agent into the neuronal cell, where the translocation domain is not a HN
CC domain of a clostridial neurotoxin and is not a fragment or derivative of
CC a HN domain of a clostridial toxin. Polypeptides of the invention are
CC useful for the treatment of a disease state associated with neuronal
CC cells. The polypeptide constructs are useful for delivering therapeutic
CC substances to neuronal cells. They are useful to treat disorders of the
CC CNS including neurodegenerative diseases, stroke, epilepsy, brain tumours
CC and infection. They are also useful in gene therapy. The present sequence
CC is C. botulinum C2 enterotoxin translocation domain with botulinum
CC neurotoxin type F (BoNT/F) binding domain used in the exemplification of
CC the invention.
XX
SQ Sequence 1092 AA;

Query Match 100.0%; Score 143; DB 22; Length 1092;
Best Local Similarity 100.0%; Pred. No. 3.6e-139;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 950 nifsntrlytgveviirkngstdisntdnfvrkndlayinvvdrdveyrlyadisiakpe 1009

Qy 61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1010 kiiklirtsnsnnslgqiivmdsignnctmnfqnnnggnigllgfhsnnlvasswyynni 1069

Qy 121 RKNTSSNGCFWSFISKEHGWQEN 143
|||||||||||||||||||||||
Db 1070 rkntssngcfwsfiskehgwqen 1092

RESULT 13
AAY77138
ID AAY77138 standard; Protein; 432 AA.
XX
AC AAY77138;
XX

DT 08-MAY-2000 (first entry)
XX
DE Synthetic botulinum neurotoxin serotype F (BoNTF) C-terminal fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype F;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
PN WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX
DR WPI; 2000-160827/14.
DR N-PSDB; AAZ87216.
XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
PS Claim 27; Page -; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAY77134-Y77139 represent
CC synthetic BoNT Hc fragments used in the present invention. The DNA
CC encoding these sequences had been optimised for codon usage for
CC expression in yeast. Note: This sequence is not given in the
CC specification, but is decoded from the BoNTF Hc DNA sequence given on
CC pages 45-46.
XX
SQ Sequence 432 AA;

Query Match 64.3%; Score 92; DB 21; Length 432;
Best Local Similarity 100.0%; Pred. No. 8.7e-87;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 ADISIAKPEKIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLV 111
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 341 adisiakpekiiklirtsnsnnslgqiivmdsignnctmnfqnnnggniglgfhsnnlv 400

Qy 112 ASSWYYNNIRKNTSSNGCFWSFISKEHGWQEN 143
||||||||||||||||||||||||||||||||
Db 401 asswyynnirkntssngcfwsfiskehgwqen 432

RESULT 14
AAW68399
ID AAW68399 standard; Protein; 448 AA.
XX
AC AAW68399;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum type F toxin C fragment.
XX
KW Antitoxin; vaccine; neurotoxin; toxin F; intoxication; immunogen;
KW botulism; BotF.
XX
OS Clostridium botulinum serotype F strain 202F (ATCC 23387).
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "N-terminal His tag"
XX
PN WO9808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA;
XX
DR WPI; 1998-230234/20.
DR N-PSDB; AAV30593.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 48; Page 364-365; 428pp; English.
XX
CC This is the amino acid sequence of the histidine-tagged C fragment
CC of Clostridium botulinum (202F strain) type F neurotoxin, encoded
CC by a DNA sequence (see AAV30593) in plasmid pETHisb. This vector
CC can be used to express BotC soluble C fragment in Escherichia
CC coli host cells, with the recombinant C fragment being purified on
CC an affinity column. The invention relates to recombinant proteins
CC derived from C. botulinum toxins, especially type B and type E
CC toxins. Methods are provided which allow for the isolation of
CC soluble recombinant proteins free of significant endotoxin
CC contamination. Preferred hosts for production of recombinant
CC proteins are E. coli, insect cells and yeast cells. The
CC recombinant toxins are used as immunogens for the production of
CC vaccines and antitoxins that are useful in the treatment of humans
CC and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 448 AA;

Query Match 16.8%; Score 24; DB 19; Length 448;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 SLGQIIVMDSIGNNCTMNFQNNNG 97
||||||||||||||||||||||||
Db 380 slgqiivmdsignnctmnfqnnng 403

RESULT 15
AAB04095
ID AAB04095 standard; Protein; 419 AA.
XX

AC AAB04095;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal sequence (serotype E).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
DR N-PSDB; AAA54489.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Disclosure; Fig 8; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 419 AA;

Query Match 5.6%; Score 8; DB 22; Length 419;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      85 GNNCTMNF 92
           ||||||||
Db     361 gnnctmnf 368
```

Search completed: August 15, 2002, 11:12:28
Job time: 320 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:14:07 ; Search time 47.36 Seconds
(without alignments)
290.135 Million cell updates/sec

Title: US-08-981-087A-4
Perfect score: 143
Sequence: 1 NIFSNTRLYTGVEVIIRKNG.....TSSNGCFWSFISKEHGWQEN 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	16.8	1274	2	I40813	neurotoxin type F
2	16	11.2	1268	2	S33411	botulinum neurotox
3	8	5.6	1252	2	S21178	botulinum neurotox
4	7	4.9	67	2	AI1594	hypothetical prote
5	7	4.9	242	2	T27999	hypothetical prote
6	7	4.9	259	2	AF1544	RNA polymerase sig
7	7	4.9	259	2	AG1186	RNA polymerase sig
8	7	4.9	264	2	G70195	pyridoxal kinase (
9	7	4.9	280	2	T47572	Machado-Joseph dis
10	7	4.9	347	2	T19989	hypothetical prote
11	7	4.9	375	2	A83636	hypothetical prote
12	7	4.9	380	2	H72374	probable phosphori
13	7	4.9	387	1	ERADN1	41K fiber protein
14	7	4.9	387	1	ERADY4	41K fiber protein
15	7	4.9	396	2	G69794	hypothetical prote
16	7	4.9	407	2	A99223	sugar phosphate nu
17	7	4.9	515	2	A42140	box B-binding fact
18	7	4.9	516	2	A44494	cAMP-responsive el
19	7	4.9	523	2	S46720	NAM8 protein - yea
20	7	4.9	543	2	T39345	probable metabolit
21	7	4.9	781	2	E64222	DNA topoisomerase
22	7	4.9	907	2	JE0176	orphan G protein-c
23	7	4.9	907	2	JG0193	G protein-coupled
24	7	4.9	1132	2	S37206	phytochrome - moss
25	7	4.9	1139	1	E64234	cytadherence-acces
26	7	4.9	1251	2	JH0256	botulinum neurotox
27	7	4.9	1303	2	C87519	hypothetical prote
28	7	4.9	1434	2	T22202	hypothetical prote
29	7	4.9	1802	2	H88444	protein C26E6.12 [
30	6	4.2	91	2	H82370	conserved hypothet
31	6	4.2	105	2	JH0239	ferredoxin precurs
32	6	4.2	108	2	B43497	hypothetical prote
33	6	4.2	114	2	T09490	hypothetical prote
34	6	4.2	117	2	B86601	CT741 hypothetical
35	6	4.2	117	2	H72022	preprotein translo
36	6	4.2	119	2	H90531	hypothetical prote
37	6	4.2	123	2	B69351	hypothetical prote
38	6	4.2	124	2	T38142	hypothetical prote
39	6	4.2	127	2	AI1748	hypothetical prote
40	6	4.2	133	2	S57492	cytochrome-c oxida
41	6	4.2	133	2	S57491	cytochrome-c oxida
42	6	4.2	133	2	S57493	cytochrome-c oxida
43	6	4.2	136	2	D70361	transcription regu
44	6	4.2	140	2	JC5003	lysozyme (EC 3.2.1
45	6	4.2	146	2	T32375	hypothetical prote

ALIGNMENTS

RESULT 1
I40813
neurotoxin type F - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C;Accession: I40813; S48108
R;East, A.K.; Richardson, P.T.; Allaway, D.; Collins, M.D.; Roberts, T.A.; Thompson,
FEMS Microbiol. Lett. 96, 225-230, 1992
A;Title: Sequence of the gene encoding type F neurotoxin of Clostridium botulinum.
A;Reference number: I40644
A;Accession: I40813
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1274 <RES>
A;Cross-references: GB:M92906; NID:g144866; PIDN:AAA23263.1; PID:g144867
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific
A;Reference number: S48103; MUID:94013372
A;Accession: S48108
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 634-1002 <CAM>
A;Cross-references: EMBL:X70816; NID:g407788; PIDN:CAA50147.1; PID:g407789
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin

Query Match 16.8%; Score 24; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 1.3e-16;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      74 SLGQIIVMDSIGNNCTMNFQNNNG 97
           ||||||||||||||||||||||||
Db    1206 SLGQIIVMDSIGNNCTMNFQNNNG 1229
```

RESULT 2
S33411
botulinum neurotoxin type F - Clostridium barati
C;Species: Clostridium barati
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999
C;Accession: S33411; S31860
R;Thompson, D.E.; Hutson, R.A.; East, A.K.; Allaway, D.; Collins, M.D.; Richardson, P
FEMS Microbiol. Lett. 108, 175-182, 1993
A;Title: Nucleotide sequence of the gene coding for Clostridium barati type F neuroto
A;Reference number: S33411; MUID:93252228
A;Accession: S33411
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1268 <THO>
A;Cross-references: EMBL:X68262; NID:g49138; PIDN:CAA48329.1; PID:g49139

```
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin


  Query Match             11.2%;  Score 16;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-08;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     127 NGCFWSFISKEHGWQE 142
           ||||||||||||||||
Db    1253 NGCFWSFISKEHGWQE 1268


RESULT   3
S21178
botulinum neurotoxin type E precursor - Clostridium botulinum
C;Species: Clostridium botulinum
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Oct-1999
C;Accession: S21178; S48107; JH0257; B35294; A60027; S18111
R;Whelan, S.M.; Elmore, M.J.; Bodsworth, N.J.; Atkinson, T.; Minton, N.P.
Eur. J. Biochem. 204, 657-667, 1992
A;Title: The complete amino acid sequence of the Clostridium botulinum type-E neurotoxin
A;Reference number: S21178; MUID:92174922
A;Accession: S21178
A;Molecule type: DNA
A;Residues: 1-1252 <WHE>
A;Cross-references: EMBL:X62683; NID:g40397; PIDN:CAA44558.1; PID:g40398
R;Campbell, K.D.; Collins, M.D.; East, A.K.
J. Clin. Microbiol. 31, 2255-2262, 1993
A;Title: Gene probes for identification of the botulinal neurotoxin gene and specific id
A;Reference number: S48103; MUID:94013372
A;Accession: S48107
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 616-982 <CAM>
A;Cross-references: EMBL:X70815; NID:g407786; PIDN:CAA50146.1; PID:g407787
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993
R;Poulet, S.; Hauser, D.; Quanz, M.; Niemann, H.; Popoff, M.R.
Biochem. Biophys. Res. Commun. 183, 107-113, 1992
A;Title: Sequences of the botulinal neurotoxin E derived from Clostridium botulinum type
A;Reference number: JH0256; MUID:92181428
A;Accession: JH0257
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-176,'R',178-197,'C',199-339,'R',341-772,'I',774-962,'FE',965-966,'R',968-1
A;Cross-references: EMBL:X62089; NID:g40393; PIDN:CAA43999.1; PID:g40394
A;Experimental source: strain Beluga
R;Binz, T.; Kurazono, H.; Wille, M.; Frevert, J.; Wernars, K.; Niemann, H.
J. Biol. Chem. 265, 9153-9158, 1990
A;Title: The complete sequence of botulinum neurotoxin type A and comparison with other
A;Reference number: A35294; MUID:90264400
A;Accession: B35294
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-176,'R',178-252 <BIN>
A;Experimental source: strain Beluga
R;Gimenez, J.A.; DasGupta, B.R.
Biochimie 72, 213-217, 1990
A;Title: Botulinum neurotoxin type E fragmented with endoproteinase Lys-C reveals the si
A;Reference number: A60027; MUID:90344918
A;Accession: A60027
A;Molecule type: protein
A;Residues: 420-427 <GIM>
A;Experimental source: strain Beluga
A;Note: this fragment was generated by proteolysis with Lys-C rather than with trypsin
C;Comment: The clostridial neurotoxins are highly potent protein toxins that inhibit neu
C;Comment: The heavy chain mediates the binding of toxin to cell receptors while the lig
C;Superfamily: tetanus toxin
C;Keywords: neurotoxin
F;1-422/Product: botulinum neurotoxin type E light chain #status predicted <LCH>
F;423-1252/Product: botulinum neurotoxin type E heavy chain #status predicted <HCH>
F;412-426/Disulfide bonds: #status predicted


  Query Match             5.6%;  Score 8;  DB 2;  Length 1252;
  Best Local Similarity   100.0%;  Pred. No. 7.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 GNNCTMNF 92
           ||||||||
Db    1194 GNNCTMNF 1201


RESULT   4
AI1594
hypothetical protein lin1298 [imported] - Listeria innocua (strain Clip11262)
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AI1594
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1594
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-67 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96529.1; PID:g16413771; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin1298


  Query Match             4.9%;  Score 7;  DB 2;  Length 67;
  Best Local Similarity   100.0%;  Pred. No. 5.6;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      39 INVVDRD 45
           |||||||
Db      31 INVVDRD 37


RESULT   5
T27999
hypothetical protein ZK795.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C;Accession: T27999
R;Percy, C.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z20453
A;Accession: T27999
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-242 <WIL>
A;Cross-references: EMBL:Z83246; PIDN:CAB05842.1; GSPDB:GN00022; CESP:ZK795.1
A;Experimental source: clone ZK795
C;Genetics:
A;Gene: CESP:ZK795.1
A;Map position: 4
A;Introns: 36/1; 59/3; 88/2; 116/1
C;Superfamily: Caenorhabditis elegans hypothetical protein ZK795.1


  Query Match             4.9%;  Score 7;  DB 2;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 18;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      56 IAKPEKI 62
           |||||||
Db     147 IAKPEKI 153
```

```
RESULT   6
AF1544
RNA polymerase sigma-37 factor (sigma-B) [imported] - Listeria innocua (strain Clip11262
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1544
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1544
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC96126.1; PID:g16413344; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: sigB
C;Superfamily: transcription sigma factor G; transcription initiation factor sigma katF


  Query Match             4.9%;  Score 7;  DB 2;  Length 259;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      98 GNIGLLG 104
           |||||||
Db      63 GNIGLLG 69


RESULT   7
AG1186
RNA polymerase sigma-37 factor (sigma-B) [imported] - Listeria monocytogenes (strain EGD
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AG1186
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AG1186
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC98973.1; PID:g16410298; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: sigB
C;Superfamily: transcription sigma factor G; transcription initiation factor sigma katF


  Query Match             4.9%;  Score 7;  DB 2;  Length 259;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      98 GNIGLLG 104
           |||||||
Db      63 GNIGLLG 69


RESULT   8
G70195
pyridoxal kinase (pdxK) homolog - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: G70195
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943
A;Accession: G70195
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-264 <KLE>
A;Cross-references: GB:AE001176; GB:AE000783; NID:g2688699; PIDN:AAC67112.1; PID:g268
A;Experimental source: strain B31


  Query Match             4.9%;  Score 7;  DB 2;  Length 264;
  Best Local Similarity   100.0%;  Pred. No. 20;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      60 EKIIKLI 66
           |||||||
Db      90 EKIIKLI 96


RESULT   9
T47572
Machado-Joseph disease MJD1a-like protein - Arabidopsis thaliana
N;Alternate names: protein F24B22.90
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47572
R;Bloecker, H.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23016
A;Accession: T47572
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-280 <BLO>
A;Cross-references: EMBL:AL132957
A;Experimental source: cultivar Columbia; BAC clone F24B22
C;Genetics:
A;Map position: 3
A;Introns: 85/3
A;Note: F24B22.90


  Query Match             4.9%;  Score 7;  DB 2;  Length 280;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     120 IRKNTSS 126
           |||||||
Db     204 IRKNTSS 210


RESULT  10
T19989
hypothetical protein C47B2.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
C;Accession: T19989
R;Kershaw, J.
submitted to the EMBL Data Library, October 1997
A;Reference number: Z19208
A;Accession: T19989
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-347 <WIL>
A;Cross-references: EMBL:Z99709; PIDN:CAB16861.1; GSPDB:GN00019; CESP:C47B2.6
A;Experimental source: clone C47B2
C;Genetics:
```

A;Gene: CESP:C47B2.6
A;Map position: 1
A;Introns: 39/1; 115/3; 258/3; 296/3
C;Superfamily: Escherichia coli UDPglucose 4-epimerase; UDPglucose 4-epimerase homology

Query Match 4.9%; Score 7; DB 2; Length 347;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 SNNLVAS 113
|||||||
Db 104 SNNLVAS 110

RESULT 11
A83636
hypothetical protein PA0063 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83636
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337
A;Accession: A83636
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-375 <STO>
A;Cross-references: GB:AE004446; GB:AE004091; NID:g9945886; PIDN:AAG03453.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0063

Query Match 4.9%; Score 7; DB 2; Length 375;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 YRLYADI 54
|||||||
Db 107 YRLYADI 113

RESULT 12
H72374
probable phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) carbon dioxide-fixation
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: H72374
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316
A;Accession: H72374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-380 <ARN>
A;Cross-references: GB:AE001723; GB:AE000512; NID:g4980953; PIDN:AAD35542.1; PID:g498096
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0447
C;Superfamily: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain; p
C;Keywords: carbon-carbon lyase; carboxy-lyase
F;15-349/Domain: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain

Query Match 4.9%; Score 7; DB 2; Length 380;
Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 41 VVDRDVE 47
|||||||
Db 349 VVDRDVE 355

RESULT 13
ERADN1
41K fiber protein - human adenovirus 41
C;Species: Mastadenovirus h41 (human adenovirus 41)
A;Note: host Homo sapiens (man)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 16-Jul-1999
C;Accession: S09217; A45352; S20695
R;Pieniazek, N.J.; Slemenda, S.B.; Pieniazek, D.; Velarde Jr., J.; Luftig, R.B.
Nucleic Acids Res. 18, 1901, 1990
A;Title: Human enteric adenovirus type 41 (Tak) contains a second fiber protein gene.
A;Reference number: S09217; MUID:90245595
A;Accession: S09217
A;Molecule type: DNA
A;Residues: 1-387 <PIE>
A;Cross-references: EMBL:X17016; NID:g58478; PIDN:CAA34882.1; PID:g58479
A;Experimental source: strain Tak
R;Kidd, A.H.; Erasmus, M.J.; Tiemessen, C.T.
Virology 179, 139-150, 1990
A;Title: Fiber sequence heterogeneity in subgroup F adenoviruses.
A;Reference number: A45352; MUID:91021015
A;Accession: A45352
A;Molecule type: DNA
A;Residues: 337-387 <KID>
A;Cross-references: GB:M60327; NID:g209932; PIDN:AAA42505.1; PID:g209933
A;Experimental source: strain FB585
R;Pienaziek, N.J.; Slemenda, S.B.; Pienazek, D.; Velarde Jr., J.; Luftig, R.B.
submitted to the EMBL Data Library, March 1990
A;Description: Characterisation of the early region E3 of the human enteric adenoviru
A;Reference number: S20688
A;Accession: S20695
A;Molecule type: DNA
A;Residues: 1-58 <PI2>
A;Cross-references: EMBL:X52198; NID:g58660; PIDN:CAA36450.1; PID:g58668
C;Superfamily: adenovirus fiber protein
C;Keywords: early protein

Query Match 4.9%; Score 7; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 NSNNSLG 76
|||||||
Db 89 NSNNSLG 95

RESULT 14
ERADY4
41K fiber protein - human adenovirus 40
C;Species: Mastadenovirus h40 (human adenovirus 40)
A;Note: host Homo sapiens (man)
C;Date: 30-Jun-1990 #sequence_revision 31-Mar-1992 #text_change 03-May-1996
C;Accession: A40048; A30336
R;Kidd, A.H.
submitted to the Protein Sequence Database, January 1992
A;Reference number: A40048
A;Accession: A40048
A;Molecule type: DNA
A;Residues: 1-387 <KID1>
R;Kidd, A.H.; Erasmus, M.J.
Virology 172, 134-144, 1989
A;Title: Sequence characterization of the adenovirus 40 fiber gene.
A;Reference number: A30336; MUID:89370295
A;Accession: A30336
A;Molecule type: DNA
A;Residues: 163-387 <KID2>

A;Cross-references: GB:M28822
C;Superfamily: adenovirus fiber protein
C;Keywords: early protein

Query Match 4.9%; Score 7; DB 1; Length 387;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 70 NSNNSLG 76
|||||||
Db 89 NSNNSLG 95

RESULT 15
G69794
hypothetical protein yerH - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: G69794
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: G69794
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-396 <KUN>
A;Cross-references: GB:Z99107; GB:AL009126; NID:g2632866; PIDN:CAB12483.1; PID:g2632977
A;Experimental source: strain 168
C;Genetics:
A;Gene: yerH
C;Superfamily: Bacillus subtilis hypothetical protein yerH

Query Match 4.9%; Score 7; DB 2; Length 396;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 20 GSTDISN 26
|||||||
Db 266 GSTDISN 272

Search completed: August 15, 2002, 11:14:08
Job time: 260 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:39 ; Search time 24.69 Seconds
(without alignments)
224.257 Million cell updates/sec

Title: US-08-981-087A-4
Perfect score: 143
Sequence: 1 NIFSNTRLYTGVEVIIRKNG.....TSSNGCFWSFISKEHGWQEN 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	24	16.8	1274	1	BXF_CLOBO	P30996 clostridium
2	7	4.9	387	1	FIB2_ADE40	P18048 human adeno
3	7	4.9	387	1	FIB2_ADE41	P16883 human adeno
4	7	4.9	515	1	CRBA_DROME	P29747 drosophila
5	7	4.9	523	1	NAM8_YEAST	Q00539 saccharomyc
6	7	4.9	781	1	PARC_MYCGE	P47446 mycoplasma
7	7	4.9	1132	1	PHY1_PHYPA	P36505 physcomitre
8	7	4.9	1139	1	HMW1_MYCGE	Q49413 mycoplasma
9	7	4.9	1250	1	BXE_CLOBO	Q00496 clostridium
10	7	4.9	1250	1	BXE_CLOBU	P30995 clostridium
11	6	4.2	64	1	Y05K_BPT4	P39238 bacteriopha
12	6	4.2	105	1	FER_SACER	P24496 saccharopol
13	6	4.2	109	1	COX1_SALTR	P29653 salmo trutt
14	6	4.2	117	1	Y884_CHLPN	Q9z722 chlamydia p
15	6	4.2	140	1	LYC_ANOGA	Q17005 anopheles g
16	6	4.2	152	1	COX1_GEOSD	P29645 geophagus s
17	6	4.2	155	1	COX1_GOMVA	P29646 gomphosus v
18	6	4.2	157	1	COX1_LEPSP	P29644 lepisosteus
19	6	4.2	157	1	COX1_SCAPL	P29654 scaphirhync
20	6	4.2	158	1	COX1_POLSP	P29650 polyodon sp
21	6	4.2	160	1	COX1_MEGAT	P29648 megalops at
22	6	4.2	161	1	COX1_POMNI	P29652 pomoxis nig
23	6	4.2	163	1	COX1_LEPOC	P29647 lepisosteus
24	6	4.2	176	1	V19R_VACCV	Q01223 vaccinia vi
25	6	4.2	178	1	RL6_HELPJ	Q9zjs6 helicobacte
26	6	4.2	178	1	RL6_HELPY	P56034 helicobacte
27	6	4.2	184	1	COX1_PANBU	P29649 pantodon bu
28	6	4.2	184	1	COX1_POLSX	P29651 polypterus
29	6	4.2	186	1	ATPD_PORPU	P51243 porphyra pu
30	6	4.2	186	1	ATPF_MESVI	Q9mut1 mesostigma
31	6	4.2	188	1	COX1_AMICA	P29643 amia calva
32	6	4.2	192	1	YS10_BORBU	P70837 borrelia bu
33	6	4.2	221	1	XYN1_COCCA	Q06562 cochliobolu
34	6	4.2	245	1	EXPR_ERWCA	Q47189 erwinia car
35	6	4.2	257	1	NT3_CHICK	P25433 gallus gall
36	6	4.2	273	1	YGF0_YEAST	P53177 saccharomyc
37	6	4.2	274	1	COX1_CHOBI	P50668 choristoneu
38	6	4.2	274	1	COX1_CHOFU	P50669 choristoneu
39	6	4.2	274	1	COX1_CHOOC	P50670 choristoneu
40	6	4.2	274	1	COX1_CHORO	P50671 choristoneu
41	6	4.2	279	1	DMSC_HAEIN	P45002 haemophilus
42	6	4.2	279	1	LPXA_CHLPN	Q9z7q4 chlamydia p
43	6	4.2	279	1	NADC_METTH	O27860 methanobact
44	6	4.2	280	1	LPXA_CHLMU	Q9pjl1 chlamydia m
45	6	4.2	280	1	LPXA_CHLTR	O84536 chlamydia t

ALIGNMENTS

RESULT 1
BXF_CLOBO
ID BXF_CLOBO STANDARD; PRT; 1274 AA.
AC P30996;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Botulinum neurotoxin type F precursor (EC 3.4.24.69) (BoNT/F)
DE (Bontoxilysin F).
GN BOTF.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23387;
RX MEDLINE=93012902; PubMed=1398040;
RA East A.K., Richardson P.T., Allaway D., Collins M.D.,
RA Roberts T.A., Thompson D.E.;
RT "Sequence of the gene encoding type F neurotoxin of Clostridium
RT botulinum.";
RL FEMS Microbiol. Lett. 75:225-230(1992).
RN [2]
RP SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=HOBBS FT10;
RX MEDLINE=94297488; PubMed=7764998;
RA East A.K., Collins M.D.;
RT "Conserved structure of genes encoding components of botulinum
RT neurotoxin complex M and the sequence of the gene coding for the
RT nontoxic component in nonproteolytic Clostridium botulinum type F.";
RL Curr. Microbiol. 29:69-77(1994).
RN [3]
RP SEQUENCE OF 634-1002 FROM N.A.
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP IDENTIFICATION OF SUBSTRATE.
RX MEDLINE=94230352; PubMed=8175689;
RA Yamasaki S., Baumeister A., Binz T., Blasi J., Link E., Cornille F.,
RA Roques B., Fykse E.M., Suedhof T.C., Jahn R., Niemann H.;
RT "Cleavage of members of the synaptobrevin/VAMP family by types D and
RT F botulinal neurotoxins and tetanus toxin.";
RL J. Biol. Chem. 269:12764-12772(1994).
CC -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 58-GLN-|-LYS-59
CC BOND OF SYNAPTOBREVINS-1 AND -2.
CC -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No

```
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M92906; AAA23263.1; -.
DR   EMBL; S73676; AAC60475.1; -.
DR   EMBL; X70820; CAA50151.1; -.
DR   EMBL; X70816; CAA50147.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   CHAIN         1    436       BOTULINUM NEUROTOXIN F, LIGHT-CHAIN.
FT   CHAIN       437   1274       BOTULINUM NEUROTOXIN F, HEAVY-CHAIN.
FT   METAL       227    227       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    228    228       BY SIMILARITY.
FT   METAL       231    231       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    429    445       INTERCHAIN (PROBABLE).
SQ   SEQUENCE   1274 AA;  146709 MW;  5B99756A7438B921 CRC64;


  Query Match             16.8%;  Score 24;  DB 1;  Length 1274;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-17;
  Matches   24;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 SLGQIIVMDSIGNNCTMNFQNNNG 97
           ||||||||||||||||||||||||
Db    1206 SLGQIIVMDSIGNNCTMNFQNNNG 1229


RESULT   2
FIB2_ADE40
ID   FIB2_ADE40     STANDARD;      PRT;   387 AA.
AC   P18048;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Fiber protein 2.
OS   Human adenovirus type 40.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=28284;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DUGAN;
RX   MEDLINE=94087748; PubMed=8263936;
RA   Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT   "The DNA sequence of adenovirus type 40.";
RL   J. Mol. Biol. 234:1308-1316(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93297140; PubMed=8517033;
RA   Kidd A.H., Chroboczek J., Cusack S., Ruigrok R.W.H.;
RT   "Adenovirus type 40 virions contain two distinct fibers.";
RL   Virology 192:73-84(1993).
RN   [3]
RP   SEQUENCE OF 167-387 FROM N.A.
RX   MEDLINE=89370295; PubMed=2773314;
RA   Kidd A.H., Erasmus M.J.;
RT   "Sequence characterization of the adenovirus 40 fiber gene.";
RL   Virology 172:134-144(1989).
CC   -!- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC       BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L19443; AAC13979.1; -.
DR   EMBL; M28822; AAA03233.1; -.
DR   PIR; A40048; ERADY4.
DR   InterPro; IPR000939; Adeno_fiber2.
DR   InterPro; IPR000978; Adeno_fiber_knob.
DR   InterPro; IPR000931; Adeno_fibre.
DR   Pfam; PF00541; adeno_fiber; 1.
DR   Pfam; PF00608; adeno_fiber2; 5.
DR   PRINTS; PR00307; ADENOVSFIBRE.
KW   Fiber protein.
FT   CONFLICT    226    226       G -> S (IN REF. 2 AND 3).
SQ   SEQUENCE   387 AA;  41346 MW;  11A3C1FCD61A3ACB CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      70 NSNNSLG 76
           |||||||
Db      89 NSNNSLG 95


RESULT   3
FIB2_ADE41
ID   FIB2_ADE41     STANDARD;      PRT;   387 AA.
AC   P16883;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Fiber protein 2.
OS   Human adenovirus type 41.
OC   Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX   NCBI_TaxID=10524;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TAK;
RX   MEDLINE=90245595; PubMed=2336370;
RA   Pieniazek N.J., Slemenda S.B., Pieniazek D., Velarde J. Jr.,
RA   Luftig R.B.;
RT   "Human enteric adenovirus type 41 (Tak) contains a second fiber
RT   protein gene.";
RL   Nucleic Acids Res. 18:1901-1901(1990).
RN   [2]
RP   SEQUENCE OF 337-387 FROM N.A.
RC   STRAIN=FB585;
RX   MEDLINE=91021015; PubMed=2219717;
RA   Kidd A.H., Erasmus M.J., Tiemessen C.T.;
RT   "Fiber sequence heterogeneity in subgroup F adenoviruses.";
RL   Virology 179:139-150(1990).
CC   -!- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC       BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X17016; CAA34882.1; -.
DR   EMBL; M60327; AAA42505.1; -.
DR   PIR; S09217; ERADN1.
DR   PIR; A45352; A45352.
DR   HSSP; P11818; 1KNB.
DR   InterPro; IPR000939; Adeno_fiber2.
DR   InterPro; IPR000978; Adeno_fiber_knob.
DR   InterPro; IPR000931; Adeno_fibre.
DR   Pfam; PF00541; adeno_fiber; 1.
DR   Pfam; PF00608; adeno_fiber2; 5.
DR   PRINTS; PR00307; ADENOVSFIBRE.
KW   Fiber protein.
SQ   SEQUENCE   387 AA;  41397 MW;  8652E785276573C7 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 387;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      70 NSNNSLG 76
           |||||||
Db      89 NSNNSLG 95


RESULT   4
CRBA_DROME
ID   CRBA_DROME     STANDARD;      PRT;   515 AA.
AC   P29747;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cyclic-AMP response element binding protein A (Box B binding factor-2)
DE   (BBF-2).
GN   CREBA OR BBBF2.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=92192458; PubMed=1532159;
RA   Abel T., Bhatt R., Maniatis T.;
RT   "A Drosophila CREB/ATF transcriptional activator binds to both fat
RT   body- and liver-specific regulatory elements.";
RL   Genes Dev. 6:466-480(1992).
CC   -!- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. BINDS TO FAT BODY-SPECIFIC
CC       ENHANCERS OF ALCOHOL DEHYDROGENASE (ADH) AND YOLK PROTEIN GENES.
CC       BBF-2 MAY PLAY A ROLE IN FAT BODY GENE EXPRESSION. IT BINDS THE
CC       CONSENSUS SEQUENCE 5'T(A/C)NACGTAN(T/G)C-3'.
CC   -!- SUBUNIT: MAY BIND DNA AS HETERODIMERS WHIT OTHER BZIP PROTEINS.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: IN ALL CELL TYPES EXAMINED.
CC   -!- DEVELOPMENTAL STAGE: PRESENT THROUGHOUT DEVELOPMENT.
CC   -!- SIMILARITY: TO OTHER BZIP PROTEINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X64429; CAA45771.1; -.
DR   PIR; S24542; S24542.
DR   PIR; A42140; A42140.
DR   TRANSFAC; T01603; -.
DR   FlyBase; FBgn0004396; CrebA.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
DR   PROSITE; PS00036; BZIP_BASIC; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DNA_BIND    442    462       BASIC MOTIF.
FT   DOMAIN      468    503       LEUCINE-ZIPPER.
SQ   SEQUENCE   515 AA;  56528 MW;  0E08FB9655200223 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 515;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 SNSNNSL 75
           |||||||
Db     264 SNSNNSL 270


RESULT   5
NAM8_YEAST
ID   NAM8_YEAST     STANDARD;      PRT;   523 AA.
AC   Q00539;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   NAM8 protein.
GN   NAM8 OR MRE2 OR YHR086W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R23/50;
RX   MEDLINE=92293106; PubMed=1603056;
RA   Ekwall K., Kermorgant M., Dujardin G., Groudinsky O.,
RA   Slonimski P.P.;
RT   "The NAM8 gene in Saccharomyces cerevisiae encodes a protein with
RT   putative RNA binding motifs and acts as a suppressor of mitochondrial
RT   splicing deficiencies when overexpressed.";
RL   Mol. Gen. Genet. 233:136-144(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Leem S.-H., Hayashi A., Ajimura M., Ogawa H.;
RL   Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- FUNCTION: ACTS AS A SUPPRESSOR OF MITOCHONDRIAL SPLICING
CC       DEFICIENCIES WHEN OVEREXPRESSED. COULD BE A NON-ESSENTIAL
CC       COMPONENT OF THE MITOCHONDRIAL SPLICING MACHINERY.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: CONTAINS 3 RNA RECOGNITION MOTIFS (RRM).
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X64763; CAA46011.1; -.
DR   EMBL; D11461; BAA02016.1; -.
DR   EMBL; U00060; AAB68928.1; -.
DR   PIR; S22439; S22439.
DR   PIR; S46720; S46720.
DR   SGD; S0001128; NAM8.
DR   InterPro; IPR000504; RRM.
DR   Pfam; PF00076; rrm; 3.
DR   SMART; SM00360; RRM; 3.
DR   PROSITE; PS50102; RRM; 3.
DR   PROSITE; PS00030; RRM_RNP_1; 1.
KW   Nuclear protein; RNA-binding; Mitochondrion; mRNA processing;
KW   Repeat.
FT   DOMAIN       54    145       RNA-BINDING (RRM) 1.
FT   DOMAIN      163    242       RNA-BINDING (RRM) 2.
FT   DOMAIN      313    385       RNA-BINDING (RRM) 3.
FT   CONFLICT    180    180       F -> L (IN REF. 1).
FT   CONFLICT    208    209       GF -> VL (IN REF. 1).
SQ   SEQUENCE   523 AA;  56972 MW;  64F198EEFB32A909 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 523;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     69 SNSNNSL 75
          |||||||
Db    148 SNSNNSL 154


RESULT   6
PARC_MYCGE
ID   PARC_MYCGE     STANDARD;      PRT;   781 AA.
AC   P47446; Q49377;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Topoisomerase IV subunit A (EC 5.99.1.-).
GN   PARC OR MG204.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 1-479 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RA   Bailey C.C., Younkins R., Huang W.M., Bott K.F.;
RL   Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: TOPOISOMERASE IV IS ESSENTIAL FOR CHROMOSOME
CC       SEGREGATION. IT HAS RELAXATION OF SUPERCOILED DNA ACTIVITY.
CC       PERFORMS THE DECATENATION EVENTS REQUIRED DURING THE REPLICATION
CC       OF A CIRCULAR DNA MOLECULE (BY SIMILARITY).
CC   -!- SUBUNIT: COMPOSED OF TWO SUBUNITS: PARC AND PARE.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC   -!- SIMILARITY: STRONG, WITH THE A SUBUNIT OF GYRASE.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U39700; AAC71422.1; -.
DR   EMBL; U25549; AAC43991.1; -.
DR   HSSP; P09097; 1AB4.
DR   TIGR; MG204; -.
DR   InterPro; IPR002205; DNA_topoisoIV.
DR   Pfam; PF00521; DNA_topoisoIV; 1.
DR   SMART; SM00434; TOP4c; 1.
KW   Topoisomerase; Isomerase; DNA-binding; Complete proteome.
FT   ACT_SITE    122    122       DNA CLEAVAGE (BY SIMILARITY).
FT   CONFLICT    261    261       P -> R (IN REF. 2).
SQ   SEQUENCE   781 AA;  88512 MW;  F14319CEE305B437 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 781;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     61 KIIKLIR 67
          |||||||
Db    388 KIIKLIR 394


RESULT   7
PHY1_PHYPA
ID   PHY1_PHYPA     STANDARD;      PRT;  1132 AA.
AC   P36505;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Phytochrome 1.
GN   PHY1.
OS   Physcomitrella patens (Moss).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
OC   Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
OX   NCBI_TaxID=3218;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94039823; PubMed=8224238;
RA   Kolukisaoglu H.U., Braun B., Martin W.F., Schneider-Poetsch H.A.W.;
RT   "Mosses do express conventional, distantly B-type-related
RT   phytochromes. Phytochrome of Physcomitrella patens (Hedw.).";
RL   FEBS Lett. 334:95-100(1993).
CC   -!- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC       ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT: THE PR FORM THAT ABSORBS
CC       MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC       ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCONVERSION OF PR IN
CC       PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES, WHEREAS
CC       RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC       RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC       GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC       BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC       PROTOCHLOROPHYLLIDE REDUCTASE, RRNA, ETC. IT ALSO CONTROLS
CC       THE EXPRESSION OF ITS OWN GENE(S) IN A NEGATIVE FEEDBACK FASHION.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- PTM: CONTAINS ONE COVALENTLY LINKED TETRAPYRROLE CHROMOPHORE.
CC   -!- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC   -!- SIMILARITY: CONTAINS 2 PAS (PER-ARNT-SIM) DIMERIZATION DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 PAS-ASSOCIATED C-TERMINAL (PAC) DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X75025; CAA52933.1; -.
DR   PIR; S37206; S37206.
DR   InterPro; IPR003018; GAF.
DR   InterPro; IPR003594; HATPase_c.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   InterPro; IPR003661; His_kinA.
DR   InterPro; IPR001610; PAC.
DR   InterPro; IPR000014; PAS.
DR   InterPro; IPR001294; Phytochrome.
DR   Pfam; PF01590; GAF; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   Pfam; PF00989; PAS; 2.
DR   Pfam; PF00360; phytochrome; 1.
DR   Pfam; PF00512; signal; 1.
DR   PRINTS; PR01033; PHYTOCHROME.
DR   SMART; SM00065; GAF; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00086; PAC; 1.
DR   SMART; SM00091; PAS; 2.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50112; PAS; 2.
DR   PROSITE; PS00245; PHYTOCHROME_1; 1.
DR   PROSITE; PS50046; PHYTOCHROME_2; 1.
KW   Transcription regulation; Photoreceptor; Phytochrome; Chromophore;
KW   Repeat.
FT   DOMAIN      610    681       PAS 1.
FT   DOMAIN      744    815       PAS 2.
FT   DOMAIN      895   1115       HISTIDINE KINASE.
FT   BINDING     321    321       CHROMOPHORE (BY SIMILARITY).
SQ   SEQUENCE   1132 AA;  125230 MW;  E1DAD4D6DC9C0D16 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 1132;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      97 GGNIGLL 103
           |||||||
Db     214 GGNIGLL 220


RESULT   8
HMW1_MYCGE
ID   HMW1_MYCGE     STANDARD;      PRT;  1139 AA.
AC   Q49413; Q49365;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cytadherence high molecular weight protein 1 (Cytadherence accessory
DE   protein 1).
GN   HMW1 OR MG312.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 721-847 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC       STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC       CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC       PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC       IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: LOCALIZES SPECIFICALLY TO THE ATTACHMENT
CC       MEMBRANE (BY SIMILARITY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U39712; AAC71534.1; -.
DR   EMBL; U02261; AAD12527.1; -.
DR   TIGR; MG312; -.
KW   Cytadherence; Structural protein; Complete proteome.
SQ   SEQUENCE   1139 AA;  130531 MW;  0011D3288C3DD856 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 1139;
  Best Local Similarity   100.0%;  Pred. No. 22;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      69 SNSNNSL 75
           |||||||
Db     450 SNSNNSL 456


RESULT   9
BXE_CLOBO
ID   BXE_CLOBO      STANDARD;      PRT;  1250 AA.
AC   Q00496;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BELUGA;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92174922; PubMed=1541280;
RA   Whelan S.M., Elmore M.J., Bodsworth N.J., Atkinson T., Minton N.P.;
RT   "The complete amino acid sequence of the Clostridium botulinum type-E
RT   neurotoxin, derived by nucleotide-sequence analysis of the encoding
RT   gene.";
RL   Eur. J. Biochem. 204:657-667(1992).
RN   [3]
RP   SEQUENCE OF 1-251 FROM N.A.
RX   MEDLINE=90264400; PubMed=2160960;
RA   Binz T., Kurazono H., Wille M., Frevert J., Wernars K., Niemann H.;
```

```
RT   "The complete sequence of botulinum neurotoxin type A and comparison
RT   with other clostridial neurotoxins.";
RL   J. Biol. Chem. 265:9153-9158(1990).
RN   [4]
RP   SEQUENCE OF 1-13.
RX   MEDLINE=85197963; PubMed=3888113;
RA   Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.;
RT   "Partial amino acid sequences of botulinum neurotoxins types B and
RT   E.";
RL   Arch. Biochem. Biophys. 238:544-548(1985).
RN   [5]
RP   SEQUENCE OF 419-426.
RX   MEDLINE=90344918; PubMed=2116911;
RA   Gimenez J.A., Dasgupta B.R.;
RT   "Botulinum neurotoxin type E fragmented with endoproteinase Lys-C
RT   reveals the site trypsin nicks and homology with tetanus
RT   neurotoxin.";
RL   Biochimie 72:213-217(1990).
RN   [6]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94063091; PubMed=8243676;
RA   Schiavo G., Santtuci A., Dasgupta B.R., Mehta P.P., Jontes J.,
RA   Benfenati F., Wilson M.C., Montecucco C.;
RT   "Botulinum neurotoxins serotypes A and E cleave SNAP-25 at distinct
RT   COOH-terminal peptide bonds.";
RL   FEBS Lett. 335:99-103(1993).
RN   [7]
RP   IDENTIFICATION OF SUBSTRATE.
RX   MEDLINE=94124495; PubMed=8294407;
RA   Binz T., Blasi J., Yamasaki S., Baumeister A., Link E., Suedhof T.C.,
RA   Jahn R., Niemann H.;
RT   "Proteolysis of SNAP-25 by types E and A botulinal neurotoxins.";
RL   J. Biol. Chem. 269:1617-1620(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CATALYZES THE HYDROLYSIS OF THE 180-ARG-|-ILE-
CC       181 BOND IN SNAP-25.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62089; CAA43999.1; -.
DR   EMBL; X62683; CAA44558.1; -.
DR   PIR; A60027; A60027.
DR   PIR; B35294; B35294.
DR   PIR; JH0257; JH0257.
DR   PIR; S08575; S08575.
DR   PIR; S18111; S18111.
DR   PIR; S21178; S21178.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    176    176       R -> G (IN REF. 2).
FT   CONFLICT    197    197       C -> S (IN REF. 2 AND 3).
FT   CONFLICT    339    339       R -> A (IN REF. 2).
FT   CONFLICT    772    772       I -> L (IN REF. 2).
FT   CONFLICT    962    963       FE -> LQ (IN REF. 2).
FT   CONFLICT    966    966       R -> A (IN REF. 2).
FT   CONFLICT   1194   1194       N -> NN (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143712 MW;  D9FCE26DDA041EB4 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      99 NIGLLGF 105
           |||||||
Db    1206 NIGLLGF 1212


RESULT  10
BXE_CLOBU
ID   BXE_CLOBU      STANDARD;      PRT;  1250 AA.
AC   P30995;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Botulinum neurotoxin type E precursor (EC 3.4.24.69) (BoNT/E)
DE   (Bontoxilysin E).
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43181, AND ATCC 43755;
RX   MEDLINE=92181428; PubMed=1543481;
RA   Poulet S., Hauser D., Quanz M., Niemann H., Popoff M.R.;
RT   "Sequences of the botulinal neurotoxin E derived from Clostridium
RT   botulinum type E (strain Beluga) and Clostridium butyricum (strains
RT   ATCC 43181 and ATCC 43755).";
RL   Biochem. Biophys. Res. Commun. 183:107-113(1992).
RN   [2]
RP   SEQUENCE OF 1-251 FROM N.A.
RC   STRAIN=BL6340;
RX   MEDLINE=91237316; PubMed=2033376;
RA   Fujii N., Kimura K., Murakami T., Indoh T., Tsuzuki K.,
RA   Yokosawa N., Yashiki T., Oguma K.;
RT   "Cloning of a DNA fragment encoding the 5'-terminus of the botulinum
RT   type E toxin gene from Clostridium butyricum strain BL6340.";
RL   J. Gen. Microbiol. 137:519-525(1991).
RN   [3]
RP   SEQUENCE OF 1-48.
RC   STRAIN=5262;
RA   Gimenez J., Foley J., Dasgupta B.R.;
RT   "Neurotoxin type E from Clostridium botulinum and C. butyricum;
RT   partial sequence and comparison.";
RL   FASEB J. 2:A1750-A1750(1988).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
```

```
CC       ENDOPEPTIDASE.
CC   -!- CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the
CC       neuroexocytosis apparatus, synaptobravins, SNAP25 or syntaxin. No
CC       detected action on small molecule substrates.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X62088; CAA43998.1; -.
DR   EMBL; X53180; CAA37321.1; -.
DR   PIR; JH0256; JH0256.
DR   PIR; S16145; S16145.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
FT   INIT_MET      0      0
FT   CHAIN         1    421       BOTULINUM NEUROTOXIN E, LIGHT-CHAIN.
FT   CHAIN       422   1250       BOTULINUM NEUROTOXIN E, HEAVY-CHAIN.
FT   METAL       211    211       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    212    212       BY SIMILARITY.
FT   METAL       215    215       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DISULFID    411    425       INTERCHAIN (PROBABLE).
FT   CONFLICT    229    229       K -> M (IN REF. 2).
SQ   SEQUENCE   1250 AA;  143265 MW;  8171B5B2C2312857 CRC64;


  Query Match             4.9%;  Score 7;  DB 1;  Length 1250;
  Best Local Similarity   100.0%;  Pred. No. 24;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     99 NIGLLGF 105
          |||||||
Db   1206 NIGLLGF 1212


RESULT  11
Y05K_BPT4
ID   Y05K_BPT4      STANDARD;      PRT;    64 AA.
AC   P39238;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 7.6 kDa protein in mobD-ri intergenic region.
GN   Y05K OR MOBD.3 OR TK.-7.
OS   Bacteriophage T4.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC   T4-like phages.
OX   NCBI_TaxID=10665;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mzhavia N., Marusich E., Djavakhishvili T., Neitzel J., Peterson S.,
RA   Awaya M., Eidermiller J., Canada D., Tracy J., Gailbreath K.,
RA   Paddison P., Anderson B., Stidham T., Blattner F., Kutter E.M.;
RT   "The 10.7 kb 'nonessential' region of bacteriophage T4 between the
RT   genes tk and nrdC: twenty new t4 genes, generally conserved among
RT   T-even phages.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
RA   Mesyanzhinov V., Ruger W., Stidham T., Thomas E.;
RT   "Bacteriophage T4 genome analysis.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U76612; AAB26966.1; -.
DR   EMBL; AF158101; AAD42595.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   64 AA;  7605 MW;  89E2AF66EE86CCE0 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 64;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     60 EKIIKL 65
          ||||||
Db      8 EKIIKL 13


RESULT  12
FER_SACER
ID   FER_SACER      STANDARD;      PRT;   105 AA.
AC   P24496;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Ferredoxin.
GN   FDXA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=WHM22;
RX   MEDLINE=91276248; PubMed=2055472;
RA   Donadio S., Hutchinson C.R.;
RT   "Cloning and characterization of the Saccharopolyspora erythraea fdxA
RT   gene encoding ferredoxin.";
RL   Gene 100:231-235(1991).
RN   [2]
RP   SEQUENCE OF 1-15.
RX   MEDLINE=88169474; PubMed=3127376;
RA   Shafiee A., Hutchinson C.R.;
RT   "Purification and reconstitution of the electron transport components
RT   for 6-deoxyerythronolide B hydroxylase, a cytochrome P-450 enzyme of
RT   macrolide antibiotic (erythromycin) biosynthesis.";
RL   J. Bacteriol. 170:1548-1553(1988).
CC   -!- FUNCTION: FERREDOXINS ARE IRON-SULFUR PROTEINS THAT TRANSFER
CC       ELECTRONS IN A WIDE VARIETY OF METABOLIC REACTIONS.
CC   -!- COFACTOR: BINDS 1 4FE-4S CLUSTER AND A 3FE-4S CLUSTER.
CC   -!- SIMILARITY: BELONGS TO THE BACTERIAL TYPE FERREDOXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M61119; AAA92023.1; -.
DR   PIR; JH0239; JH0239.
DR   HSSP; Q45560; 1BD6.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR000813; 7Fe_ferredoxin.
DR   Pfam; PF00037; fer4; 1.
DR   PRINTS; PR00353; 4FE4SFRDOXIN.
DR   PRINTS; PR00354; 7FE8SFRDOXIN.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW   Electron transport; Iron-sulfur; Duplication; 4Fe-4S; 3Fe-4S.
FT   INIT_MET      0      0
FT   METAL         8      8       IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).
FT   METAL        16     16       IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).
FT   METAL        20     20       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        39     39       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        42     42       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        45     45       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        49     49       IRON-SULFUR 1 (3FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   105 AA;  11407 MW;  F42D85AC36406683 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 105;
  Best Local Similarity   100.0%;  Pred. No. 32;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     42 VDRDVE 47
          ||||||
Db     90 VDRDVE 95


RESULT  13
COX1_SALTR
ID   COX1_SALTR     STANDARD;      PRT;   109 AA.
AC   P29653;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Fragment).
GN   COXI OR COI.
OS   Salmo trutta (Brown trout).
OG   Mitochondrion.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC   Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX   NCBI_TaxID=8032;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92130804; PubMed=1663569;
RA   Normark B.B., McCune A.R., Harrison R.G.;
RT   "Phylogenetic relationships of neopterygian fishes, inferred from
RT   mitochondrial DNA sequences.";
RL   Mol. Biol. Evol. 8:819-834(1991).
CC   -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC       CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC       3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC       CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC       CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC       AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC       AND COPPER B.
CC   -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome
CC       c + 2 H(2)O.
CC   -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC       INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M64917; AAB01484.1; -.
DR   HSSP; P00396; 1OCC.
DR   InterPro; IPR000883; COX1.
DR   Pfam; PF00115; COX1; 1.
DR   PROSITE; PS00077; COX1; 1.
KW   Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
KW   Respiratory chain; Inner membrane.
FT   NON_TER       1      1
FT   METAL         6      6       COPPER B (PROBABLE).
FT   METAL        10     10       COPPER B (PROBABLE).
FT   METAL        56     56       COPPER B (PROBABLE).
FT   METAL        57     57       COPPER B (PROBABLE).
FT   NON_TER     109    109
SQ   SEQUENCE   109 AA;  12251 MW;  0A513F2CE5B85C25 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 109;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    100 IGLLGF 105
          ||||||
Db     46 IGLLGF 51


RESULT  14
Y884_CHLPN
ID   Y884_CHLPN     STANDARD;      PRT;   117 AA.
AC   Q9Z722; Q9JQ92;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein CPn0884/CP0982/CPj0884.
GN   CPN0884 OR CP0982 OR CPJ0884.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SIMILARITY: BELONGS TO THE UPF0092 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE001669; AAD19022.1; -.
DR   EMBL; AE002256; AAF38761.1; -.
DR   EMBL; AP002548; BAA99092.1; -.
DR   TIGR; CP0982; -.
DR   InterPro; IPR003849; DUF219.
DR   Pfam; PF02699; DUF219; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     30     50       POTENTIAL.
SQ   SEQUENCE   117 AA;  13152 MW;  45D3D3AC8B9E11A2 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 117;
  Best Local Similarity   100.0%;  Pred. No. 36;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      32 RKNDLA 37
           ||||||
Db      63 RKNDLA 68


RESULT  15
LYC_ANOGA
ID   LYC_ANOGA      STANDARD;      PRT;   140 AA.
AC   Q17005;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Lysozyme precursor (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase).
OS   Anopheles gambiae (African malaria mosquito).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC   Anopheles.
OX   NCBI_TaxID=7165;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97045819; PubMed=8890741;
RA   Kang D., Romans P., Lee J.Y.;
RT   "Analysis of a lysozyme gene from the malaria vector mosquito,
RT   Anopheles gambiae.";
RL   Gene 174:239-244(1996).
CC   -!- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE
CC       IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
CC       MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC       acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC       heteropolymers of the prokaryotes cell walls.
CC   -!- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U28809; AAC47326.1; -.
DR   HSSP; P00698; 1AT5.
DR   InterPro; IPR001916; Lactalbmn_lysozyme.
DR   Pfam; PF00062; lys; 1.
DR   PRINTS; PR00135; LYZLACT.
DR   SMART; SM00263; LYZ1; 1.
DR   PROSITE; PS00128; LACTALBUMIN_LYSOZYME; 1.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    140       LYSOZYME.
FT   DISULFID     26    139       BY SIMILARITY.
FT   DISULFID     47    128       BY SIMILARITY.
FT   DISULFID     81     94       BY SIMILARITY.
FT   DISULFID     90    108       BY SIMILARITY.
FT   ACT_SITE     52     52       BY SIMILARITY.
FT   ACT_SITE     69     69       BY SIMILARITY.
SQ   SEQUENCE   140 AA;  15398 MW;  93AD614699216C28 CRC64;


  Query Match             4.2%;  Score 6;  DB 1;  Length 140;
  Best Local Similarity   100.0%;  Pred. No. 42;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 KNGSTD 23
           ||||||
Db      64 KNGSTD 69


Search completed: August 15, 2002, 11:24:40
Job time: 687 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:06 ; Search time 80.39 Seconds
(without alignments)
309.880 Million cell updates/sec

Title: US-08-981-087A-2
Perfect score: 144
Sequence: 1 SYTNDKILILYFNKLYKKIK.....LNYNKIIWTLQDTAGNNQKL 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144	100.0	1278	2	Q57236	Q57236 clostridium
2	26	18.1	1280	2	Q9ZAJ5	Q9zaj5 clostridium
3	15	10.4	1268	2	Q45851	Q45851 clostridium
4	11	7.6	367	2	Q45862	Q45862 clostridium
5	11	7.6	367	2	Q45861	Q45861 clostridium
6	11	7.6	1251	2	Q9K395	Q9k395 clostridium
7	11	7.6	1255	2	Q9FAR6	Q9far6 clostridium
8	9	6.2	1296	2	Q45894	Q45894 clostridium
9	8	5.6	361	2	Q45846	Q45846 clostridium
10	8	5.6	361	2	Q45848	Q45848 clostridium
11	8	5.6	441	2	Q9X708	Q9x708 clostridium
12	8	5.6	1072	16	Q9CF64	Q9cf64 lactococcus
13	8	5.6	1291	2	Q9ZAJ8	Q9zaj8 clostridium
14	8	5.6	1291	2	Q08077	Q08077 clostridium
15	8	5.6	1291	2	Q93G71	Q93g71 clostridium
16	8	5.6	1291	2	Q933K0	Q933k0 clostridium
17	7	4.9	53	13	Q90WI9	Q90wi9 gallus gall
18	7	4.9	111	13	Q90X55	Q90x55 gallus gall
19	7	4.9	132	13	Q90X56	Q90x56 gallus gall
20	7	4.9	169	9	Q9G0E2	Q9g0e2 lactococcus
21	7	4.9	177	16	O51665	O51665 borrelia bu
22	7	4.9	209	17	Q97VC3	Q97vc3 sulfolobus
23	7	4.9	241	17	Q97UQ7	Q97uq7 sulfolobus
24	7	4.9	244	12	Q9YVU6	Q9yvu6 melanoplus
25	7	4.9	247	2	Q9KWM3	Q9kwm3 staphylococ
26	7	4.9	261	10	Q9SIF3	Q9sif3 arabidopsis
27	7	4.9	330	16	Q99WV1	Q99wv1 staphylococ
28	7	4.9	448	2	Q9AKD7	Q9akd7 rickettsia
29	7	4.9	458	5	Q9NAP3	Q9nap3 caenorhabdi
30	7	4.9	484	5	Q22967	Q22967 caenorhabdi
31	7	4.9	648	3	Q9C469	Q9c469 schizosacch
32	7	4.9	737	16	O51274	O51274 borrelia bu
33	7	4.9	842	2	Q93U59	Q93u59 candidatus
34	7	4.9	865	12	Q913V4	Q913v4 porcine gam
35	7	4.9	885	5	Q9BHY8	Q9bhy8 leishmania
36	7	4.9	890	10	Q93ZT8	Q93zt8 arabidopsis
37	7	4.9	1010	5	O18346	O18346 drosophila
38	7	4.9	1037	5	Q9N9L3	Q9n913 leishmania
39	7	4.9	1116	5	Q19643	Q19643 caenorhabdi
40	7	4.9	1711	5	O96108	O96108 plasmodium
41	7	4.9	2292	12	Q66763	Q66763 encephalomy
42	7	4.9	2292	12	Q89272	Q89272 encephalomy
43	6	4.2	52	2	Q9R5H1	Q9r5h1 clostridium
44	6	4.2	56	10	Q38975	Q38975 arabidopsis
45	6	4.2	57	10	Q38974	Q38974 arabidopsis

ALIGNMENTS

RESULT 1
Q57236
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).

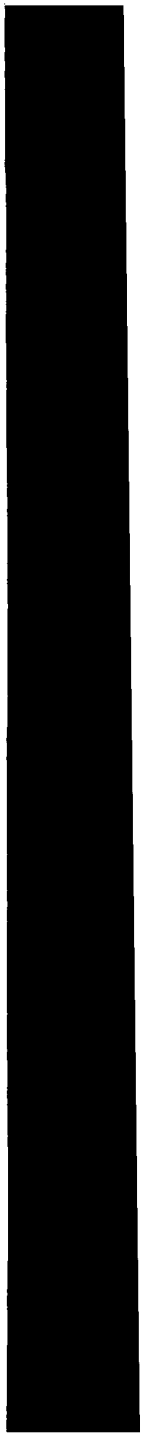
```
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;


  Query Match             100.0%;  Score 144;  DB 2;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-135;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     848 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 907

Qy      61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     908 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 967

Qy     121 WKISLNYNKIIWTLQDTAGNNQKL 144
           ||||||||||||||||||||||||
Db     968 WKISLNYNKIIWTLQDTAGNNQKL 991


RESULT   2
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BONT PROTEIN.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281 (ATCC 43757);
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;


  Query Match             18.1%;  Score 26;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-17;
  Matches   26;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 SILDMRYENNKFIDISGYGSNISING 48
           ||||||||||||||||||||||||||
Db     870 SILDMRYENNKFIDISGYGSNISING 895


RESULT   3
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROTOXIN TYPE F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;


  Query Match             10.4%;  Score 15;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      51 YIYSTNRNQFGIYSS 65
           |||||||||||||||
Db     889 YIYSTNRNQFGIYSS 903


RESULT   4
Q45862
ID   Q45862      PRELIMINARY;      PRT;   367 AA.
AC   Q45862;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE E, HAZEN 36208 (ATCC 9564);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70815; CAA50146.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42854 MW;  0810595B3A865570 CRC64;
```

```
  Query Match             7.6%;  Score 11;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 0.0072;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||| 
Db     299 NFSISFWVRIP 309


RESULT   5
Q45861
ID   Q45861      PRELIMINARY;      PRT;   367 AA.
AC   Q45861;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE E, VH (DOLMAN);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70818; CAA50149.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42902 MW;  346A610C2FF70262 CRC64;


  Query Match             7.6%;  Score 11;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 0.0072;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||| 
Db     299 NFSISFWVRIP 309


RESULT   6
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155 (KZ 1885);
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
```

```
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;


  Query Match             7.6%;  Score 11;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||| 
Db     914 NFSISFWVRIP 924


RESULT   7
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;


  Query Match             7.6%;  Score 11;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 0.018;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      86 NFSISFWVRIP 96
           ||||||||||| 
Db     917 NFSISFWVRIP 927


RESULT   8
Q45894
ID   Q45894      PRELIMINARY;      PRT;  1296 AA.
AC   Q45894; P77780;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE A (TYPE A NEUROTOXIN).
GN   BONT OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KYOTO-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [3]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1296 AA;  149410 MW;  6F12E7BF28ED69D1 CRC64;


  Query Match             6.2%;  Score 9;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 1.9;
  Matches    9;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     103 NLNNEYTII 111
           |||||||||
Db     957 NLNNEYTII 965


RESULT   9
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
```

```
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    117 NNSGWKIS 124
          ||||||||
Db    325 NNSGWKIS 332


RESULT  10
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE B, NON-PROTEOLYTIC EKLUND 2B (COLWORTH 229);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 7.1;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    117 NNSGWKIS 124
          ||||||||
Db    325 NNSGWKIS 332


RESULT  11
Q9X708
ID   Q9X708      PRELIMINARY;      PRT;   441 AA.
AC   Q9X708;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99343691; PubMed=10413679;
RA   Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA   Schiavo G.;
RT   "Functional characterisation of tetanus and botulinum neurotoxins
RT   binding domains.";
RL   J. Cell Sci. 112:2715-2724(1999).
DR   EMBL; AJ242628; CAB43706.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     441    441
SQ   SEQUENCE   441 AA;  52772 MW;  721D0B468E8C95A4 CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 441;
  Best Local Similarity   100.0%;  Pred. No. 8.3;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    117 NNSGWKIS 124
          ||||||||
Db    116 NNSGWKIS 123


RESULT  12
Q9CF64
ID   Q9CF64      PRELIMINARY;      PRT;  1072 AA.
AC   Q9CF64;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   UNKNOWN PROTEIN.
GN   YQFG.
OS   Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC   Lactococcus.
OX   NCBI_TaxID=1360;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IL1403;
RX   MEDLINE=21235186; PubMed=11337471;
RA   Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
RA   Weissenbach J., Ehrlich S.D., Sorokin A.;
RT   "The complete genome sequence of the lactic acid bacterium Lactococcus
RT   lactis ssp. lactis IL1403.";
RL   Genome Res. 11:731-753(2001).
DR   EMBL; AE006392; AAK05715.1; -.
KW   Complete proteome.
SQ   SEQUENCE   1072 AA;  113056 MW;  464446E2656CAA08 CRC64;


  Query Match             5.6%;  Score 8;  DB 16;  Length 1072;
  Best Local Similarity   100.0%;  Pred. No. 16;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     64 SSKPSEVN 71
          ||||||||
Db    712 SSKPSEVN 719


RESULT  13
Q9ZAJ8
ID   Q9ZAJ8      PRELIMINARY;      PRT;  1291 AA.
AC   Q9ZAJ8;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BONT PROTEIN.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=CDC 3281 (ATCC 43757);
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13630; CAA73968.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1291 AA;  150840 MW;  E4D3B0E46AB2E735 CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     958 NNSGWKIS 965


RESULT  14
Q08077
ID   Q08077      PRELIMINARY;      PRT;  1291 AA.
AC   Q08077;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (EC 3.4.24.-) (BONT/B).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EKLUND 17B ATCC25765;
RX   MEDLINE=94122659; PubMed=7764370;
RA   Hutson R.A., Collins M.D., East A.K., Thompson D.E.;
RT   "Nucleotide sequence of the gene coding for non-proteolytic
RT   Clostridium botulinum type B neurotoxin: comparison with other
RT   clostridial neurotoxins.";
RL   Curr. Microbiol. 28:101-110(1994).
CC   -!- FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER
CC       RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED
CC       AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD
CC       WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT
CC       INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC
CC       ENDOPEPTIDASE THAT CLEAVES SYNAPTOBREVIN-2.
CC   -!- SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A A
CC       HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY,
CC       WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL
CC       FORMATION AND TOXIN BINDING, RESPECTIVELY.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF
CC       BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.
CC   -!- SIMILARITY: HIGH WITH OTHER BOTULINUM NEUROTOXINS AND WITH TETANUS
CC       NEUROTOXIN.
CC   -!- SIMILARITY: TO OTHER ZINC METALLOPROTEINASES IN THE ACTIVE SITE
CC       REGION.
DR   EMBL; X71343; CAA50482.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc.
SQ   SEQUENCE   1291 AA;  150513 MW;  71BCAFE23D69FAAA CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     958 NNSGWKIS 965


RESULT  15
Q93G71
ID   Q93G71      PRELIMINARY;      PRT;  1291 AA.
AC   Q93G71;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROTOXIN TYPE B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=1436;
RA   Kirma N., Ferreira J.L., Baumstark B.R.;
RT   "Characterization of six type A strains of Clostridium botulinum that
RT   contain type B toxin gene sequences.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF295926; AAK97132.1; -.
SQ   SEQUENCE   1291 AA;  150824 MW;  D7CA07BAE2EB8CD2 CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 1291;
  Best Local Similarity   100.0%;  Pred. No. 19;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     117 NNSGWKIS 124
           ||||||||
Db     958 NNSGWKIS 965


Search completed: August 15, 2002, 11:24:07
Job time: 694 sec
```



GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:07 ; Search time 80.39 Seconds
(without alignments)
309.880 Million cell updates/sec

Title: US-08-981-087A-3
Perfect score: 144
Sequence: 1 VFNYTQMISISDYINKWIFV.....ITQNSNFLNINQQRGVYQKP 144

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	144	100.0	1278	2	Q57236	Q57236 clostridium
2	31	21.5	1280	2	Q9ZAJ5	Q9zaj5 clostridium
3	22	15.3	1268	2	Q45851	Q45851 clostridium
4	15	10.4	1251	2	Q9K395	Q9k395 clostridium
5	15	10.4	1255	2	Q9FAR6	Q9far6 clostridium
6	11	7.6	1296	2	Q45894	Q45894 clostridium
7	8	5.6	540	16	Q9A9Z8	Q9a9z8 caulobacter
8	7	4.9	131	4	Q9BUA6	Q9bua6 homo sapien
9	7	4.9	154	11	Q62080	Q62080 mus musculu
10	7	4.9	166	4	O95023	O95023 homo sapien
11	7	4.9	182	15	Q85641	Q85641 moloney mur
12	7	4.9	187	15	Q85640	Q85640 murine leuk
13	7	4.9	187	15	Q83401	Q83401 moloney mur
14	7	4.9	200	16	Q9PMH8	Q9pmh8 campylobact
15	7	4.9	202	11	Q62082	Q62082 mus musculu
16	7	4.9	203	17	Q979N7	Q979n7 thermoplasm
17	7	4.9	231	2	Q46032	Q46032 clostridium
18	7	4.9	232	2	Q46027	Q46027 clostridium
19	7	4.9	263	2	Q93LJ8	Q93lj8 neisseria m
20	7	4.9	263	2	Q93LJ6	Q93lj6 neisseria l
21	7	4.9	267	16	Q9JXT3	Q9jxt3 neisseria m
22	7	4.9	307	16	Q9HXW8	Q9hxw8 pseudomonas
23	7	4.9	358	12	Q9WR17	Q9wr17 african cas
24	7	4.9	358	12	Q9JEA9	Q9jea9 cassava gem
25	7	4.9	358	12	Q9JEA2	Q9jea2 cassava gem
26	7	4.9	379	17	O29504	O29504 archaeoglob
27	7	4.9	436	10	Q9FVZ5	Q9fvz5 oryza sativ
28	7	4.9	441	2	Q9X708	Q9x708 clostridium
29	7	4.9	450	12	Q9EMU0	Q9emu0 amsacta moo
30	7	4.9	461	16	Q97E21	Q97e21 clostridium
31	7	4.9	534	5	Q9U211	Q9u211 caenorhabdi
32	7	4.9	636	15	Q85506	Q85506 murine leuk
33	7	4.9	669	15	Q9YYS3	Q9yys3 murine leuk
34	7	4.9	814	10	O22695	O22695 arabidopsis
35	7	4.9	846	10	Q9FP99	Q9fp99 oryza sativ
36	7	4.9	1023	5	O62398	O62398 caenorhabdi
37	7	4.9	1036	11	Q91YD5	Q91yd5 mus musculu
38	7	4.9	1270	5	Q19736	Q19736 caenorhabdi
39	7	4.9	1272	12	O10243	O10243 clover yell
40	7	4.9	1275	12	Q9QTG7	Q9qtg7 clostridium
41	7	4.9	1280	2	Q9LBS7	Q9lbs7 clostridium
42	7	4.9	1280	2	Q45849	Q45849 clostridium
43	7	4.9	1291	2	Q9ZAJ8	Q9zaj8 clostridium
44	7	4.9	1291	2	Q08077	Q08077 clostridium
45	7	4.9	1291	2	Q93G71	Q93g71 clostridium

ALIGNMENTS

RESULT 1
Q57236
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).

```
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;


  Query Match           100.0%;  Score 144;  DB 2;  Length 1278;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-138;
  Matches  144;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992 VFNYTQMISISDYINKWIFVTITNNRLGNSRIYINGNLIDEKSISNLGDIHVSDNILFKI 1051

Qy     61 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1052 VGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDPSILKDFWGNYLLYNKRYYLLNLLR 1111

Qy    121 TDKSITQNSNFLNINQQRGVYQKP 144
          ||||||||||||||||||||||||
Db   1112 TDKSITQNSNFLNINQQRGVYQKP 1135


RESULT   2
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BONT PROTEIN.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281 (ATCC 43757);
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;


  Query Match           21.5%;  Score 31;  DB 2;  Length 1280;
  Best Local Similarity  100.0%;  Pred. No. 5.3e-23;
  Matches   31;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      9 SISDYINKWIFVTITNNRLGNSRIYINGNLI 39
          |||||||||||||||||||||||||||||||
Db   1003 SISDYINKWIFVTITNNRLGNSRIYINGNLI 1033


RESULT   3
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROTOXIN TYPE F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;


  Query Match           15.3%;  Score 22;  DB 2;  Length 1268;
  Best Local Similarity  100.0%;  Pred. No. 8.4e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     54 DNILFKIVGCNDTRYVGIRYFK 75
          ||||||||||||||||||||||
Db   1036 DNILFKIVGCNDTRYVGIRYFK 1057


RESULT   4
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155 (KZ 1885);
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;


  Query Match             10.4%;  Score 15;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           ||||||||||||||
Db     983 ISDYINKWIFVTITN 997


RESULT   5
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;


  Query Match             10.4%;  Score 15;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 ISDYINKWIFVTITN 24
           ||||||||||||||
Db     986 ISDYINKWIFVTITN 1000


RESULT   6
Q45894
ID   Q45894      PRELIMINARY;      PRT;  1296 AA.
AC   Q45894; P77780;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE A (TYPE A NEUROTOXIN).
GN   BONT OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
```

```
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KYOTO-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [3]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1296 AA;  149410 MW;  6F12E7BF28ED69D1 CRC64;


  Query Match             7.6%;  Score 11;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.015;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 WIFVTITNNRL 27
           ||||||||||| 
Db    1014 WIFVTITNNRL 1024


RESULT   7
Q9A9Z8
ID   Q9A9Z8      PRELIMINARY;      PRT;   540 AA.
AC   Q9A9Z8;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN CC0813.
GN   CC0813.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=69394;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005758; AAK22798.1; -.
DR   TIGR; CC0813; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   540 AA;  59648 MW;  72BC45442BEF99FD CRC64;


  Query Match             5.6%;  Score 8;  DB 16;  Length 540;
  Best Local Similarity   100.0%;  Pred. No. 8.7;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      94 PDPSILKD 101
           ||||||||
Db      71 PDPSILKD 78


RESULT   8
Q9BUA6
ID   Q9BUA6      PRELIMINARY;      PRT;   131 AA.
AC   Q9BUA6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SIMILAR TO MYOSIN LIGHT CHAIN 2, PRECURSOR LYMPHOCYTE-SPECIFIC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA.;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC002778; AAH02778.1; -.
DR   HSSP; P13543; 1SCM.
DR   InterPro; IPR002048; EF-hand.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
SQ   SEQUENCE   131 AA;  14930 MW;  336C55EE3C70C07A CRC64;


  Query Match             4.9%;  Score 7;  DB 4;  Length 131;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db      68 FKVFDTE 74


RESULT   9
Q62080
ID   Q62080      PRELIMINARY;      PRT;   154 AA.
AC   Q62080;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MYOSIN LIGHT CHAIN 2 (FRAGMENT).
GN   MYLC2PL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92331628; PubMed=1628631;
RA   Oltz E.M., Yancopoulos G.D., Morrow M.A., Rolink A., Lee G., Wong F.,
RA   Kaplan K., Gillis S., Melchers F., Alt F.W.;
RT   "A novel regulatory myosin light chain gene distinguishes pre-B cell
RT   subsets and is IL-7 inducible.";
RL   EMBO J. 11:2759-2767(1992).
```

```
DR   EMBL; X65979; CAA46794.1; -.
DR   HSSP; P13543; 1SCM.
DR   MGD; MGI:1891705; Mylc2pl.
FT   NON_TER       1      1
SQ   SEQUENCE   154 AA;  17439 MW;  085265DBAE42912F CRC64;


  Query Match             4.9%;  Score 7;  DB 11;  Length 154;
  Best Local Similarity   100.0%;  Pred. No. 33;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db      91 FKVFDTE 97


RESULT  10
O95023
ID   O95023      PRELIMINARY;      PRT;   166 AA.
AC   O95023;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   WUGSC:H_DJ1059M17.2 PROTEIN (FRAGMENT).
GN   WUGSC:H_DJ1059M17.2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Geisel C., Kalicki J., Gibson A.;
RT   "The sequence of Homo sapiens PAC clone RP5-1059M17.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Waterston R.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR   EMBL; AC004953; AAD08850.1; -.
DR   HSSP; P13543; 1SCM.
DR   InterPro; IPR002048; EF-hand.
DR   Pfam; PF00036; efhand; 2.
DR   SMART; SM00054; EFh; 2.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW   Calcium-binding.
FT   NON_TER       1      1
SQ   SEQUENCE   166 AA;  18917 MW;  BCD43CB94A931605 CRC64;


  Query Match             4.9%;  Score 7;  DB 4;  Length 166;
  Best Local Similarity   100.0%;  Pred. No. 35;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db     103 FKVFDTE 109


RESULT  11
Q85641
ID   Q85641      PRELIMINARY;      PRT;   182 AA.
AC   Q85641;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   3' END OF THE GENOME OF MOLONEY MURINE LEUKEMIA VIRUS (CODES FOR THE
DE   ENV GENE) (FRAGMENT).
OS   Moloney murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11801;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81013872; PubMed=6251454;
RA   Sutcliffe J.G., Shinnick T.M., Verma I.M., Lerner R.A.;
RT   "Nucleotide sequence of Moloney leukemia virus: 3' end reveals details
RT   of replications, analogy to bacterial transposons, and an unexpected
RT   gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).
DR   EMBL; V01178; CAA24501.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; ENV_polyprotein; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   182 AA;  20384 MW;  9212B2B384E97328 CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 182;
  Best Local Similarity   100.0%;  Pred. No. 37;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 EKSISNL 47
           |||||||
Db      37 EKSISNL 43


RESULT  12
Q85640
ID   Q85640      PRELIMINARY;      PRT;   187 AA.
AC   Q85640;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MOLONEY MURINE LEUKEMIA VIRUS (M-MULV) TRANSPOSON GENES AND
DE   REPLICATING DETAILS (FRAGMENT).
OS   Murine leukemia virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11786;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81259595; PubMed=6266762;
RA   Sutcliffe J.G., Shinnick T.M., Lerner R.A.;
RT   "Moloney murine leukemia virus is a transposon: Nucleotide sequence
RT   analysis identifies genes and replication details.";
RL   Cold Spring Harb. Symp. Quant. Biol. 45:707-710(1981).
DR   EMBL; M12997; AAA46529.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; ENV_polyprotein; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   187 AA;  20841 MW;  C0832DDFE4BD7CA5 CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 187;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 EKSISNL 47
           |||||||
Db      42 EKSISNL 48


RESULT  13
Q83401
ID   Q83401      PRELIMINARY;      PRT;   187 AA.
AC   Q83401;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   POLYPROTEIN PRECURSOR (FRAGMENT).
OS   Moloney murine sarcoma virus.
OC   Viruses; Retroid viruses; Retroviridae; Gammaretrovirus.
OX   NCBI_TaxID=11809;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=81052384; PubMed=6159543;
```

```
RA   Sutcliffe J.G., Shinnick T.M., Green N., Liu F.T., Niman H.L.,
RA   Lerner R.A.;
RT   "Chemical synthesis of a polypeptide predicted from nucleotide
RT   sequence allows detection of a new retroviral gene product.";
RL   Nature 287:801-805(1980).
RN   [2]
RP   SEQUENCE OF 6-187 FROM N.A.
RX   MEDLINE=81013872; PubMed=6251454;
RA   Sutcliffe J.G., Shinnick T.M., Verma I.M., Lerner R.A.;
RT   "Nucleotide sequence of Moloney leukemia virus: 3' end reveals details
RT   of replications, analogy to bacterial transposons, and an unexpected
RT   gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 77:3302-3306(1980).
DR   EMBL; J02261; AAA51623.1; -.
DR   HSSP; P03385; 1MOF.
DR   InterPro; IPR002050; Env_polyprotein.
DR   Pfam; PF00429; ENV_polyprotein; 1.
FT   NON_TER       1      1
FT   CHAIN         1     94       POTENTIAL.
FT   CHAIN        95    187       POTENTIAL.
SQ   SEQUENCE   187 AA;  20841 MW;  C0832DDFE4BD7CA5 CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 187;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      41 EKSISNL 47
           |||||||
Db      42 EKSISNL 48


RESULT  14
Q9PMH8
ID   Q9PMH8      PRELIMINARY;      PRT;   200 AA.
AC   Q9PMH8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE MEMBRANE PROTEIN.
GN   CJ1484C.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC   Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
DR   EMBL; AL139078; CAB73906.1; -.
KW   Complete proteome.
SQ   SEQUENCE   200 AA;  22387 MW;  753C07F428DA45BE CRC64;


  Query Match             4.9%;  Score 7;  DB 16;  Length 200;
  Best Local Similarity   100.0%;  Pred. No. 40;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      83 KTEIETL 89
           |||||||
Db      66 KTEIETL 72


RESULT  15
Q62082
```

```
ID   Q62082      PRELIMINARY;      PRT;   202 AA.
AC   Q62082;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MYOSIN LIGHT CHAIN 2.
GN   MYLC2PL.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92331628; PubMed=1628631;
RA   Oltz E.M., Yancopoulos G.D., Morrow M.A., Rolink A., Lee G., Wong F.,
RA   Kaplan K., Gillis S., Melchers F., Alt F.W.;
RT   "A novel regulatory myosin light chain gene distinguishes pre-B cell
RT   subsets and is IL-7 inducible.";
RL   EMBO J. 11:2759-2767(1992).
CC   -!- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR   EMBL; X65981; CAA46796.1; -.
DR   HSSP; P13543; 1SCM.
DR   MGD; MGI:1891705; Mylc2pl.
DR   InterPro; IPR002048; EF-hand.
DR   Pfam; PF00036; efhand; 2.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW   Calcium-binding.
SQ   SEQUENCE   202 AA;  22511 MW;  97C976AD9E1A90BD CRC64;


  Query Match             4.9%;  Score 7;  DB 11;  Length 202;
  Best Local Similarity   100.0%;  Pred. No. 41;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      74 FKVFDTE 80
           |||||||
Db     139 FKVFDTE 145


Search completed: August 15, 2002, 11:24:08
Job time: 695 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:24:08 ; Search time 80.39 Seconds
(without alignments)
307.728 Million cell updates/sec

Title: US-08-981-087A-4
Perfect score: 143
Sequence: 1 NIFSNTRLYTGVEVIIRKNG.....TSSNGCFWSFISKEHGWQEN 143

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	143	100.0	1278	2	Q57236	Q57236 clostridium
2	22	15.4	1280	2	Q9ZAJ5	Q9zaj5 clostridium
3	16	11.2	1268	2	Q45851	Q45851 clostridium
4	8	5.6	96	15	Q994N3	Q994n3 human immun
5	8	5.6	96	15	Q99BN5	Q99bn5 human immun
6	8	5.6	731	5	Q9BPL0	Q9bpl0 schistosoma
7	8	5.6	1251	2	Q9K395	Q9k395 clostridium
8	8	5.6	1255	2	Q9FAR6	Q9far6 clostridium
9	7	4.9	67	16	Q92C93	Q92c93 listeria in
10	7	4.9	96	15	O42077	O42077 human immun
11	7	4.9	96	15	O89653	O89653 human immun
12	7	4.9	96	15	O89654	O89654 human immun
13	7	4.9	96	15	O89670	O89670 human immun
14	7	4.9	96	15	O89677	O89677 human immun
15	7	4.9	96	15	O89678	O89678 human immun
16	7	4.9	96	15	O89679	O89679 human immun
17	7	4.9	96	15	Q90QQ8	Q90qq8 human immun
18	7	4.9	96	15	Q90DZ9	Q90dz9 human immun
19	7	4.9	180	15	Q9QRP6	Q9qrp6 human immun
20	7	4.9	242	3	Q9P829	Q9p829 candida alb
21	7	4.9	242	5	O62519	O62519 caenorhabdi
22	7	4.9	249	8	Q95CI9	Q95ci9 sarcopera r
23	7	4.9	252	8	Q95CI7	Q95ci7 schwartzia
24	7	4.9	252	8	Q95CI6	Q95ci6 souroubea e
25	7	4.9	255	8	Q95CJ2	Q95cj2 marcgravias
26	7	4.9	255	8	Q95CJ0	Q95cj0 ruyschia ph
27	7	4.9	255	8	Q95CI8	Q95ci8 sarcopera s
28	7	4.9	255	8	Q95CI5	Q95ci5 souroubea l
29	7	4.9	255	8	Q95CI4	Q95ci4 souroubea s
30	7	4.9	259	2	Q9R9A4	Q9r9a4 listeria mo
31	7	4.9	259	2	O85017	O85017 listeria in
32	7	4.9	259	16	Q92DC1	Q92dc1 listeria in
33	7	4.9	264	16	O51709	O51709 borrelia bu
34	7	4.9	266	2	O85743	O85743 listeria mo
35	7	4.9	280	10	Q9M391	Q9m391 arabidopsis
36	7	4.9	347	5	O62107	O62107 caenorhabdi
37	7	4.9	375	16	Q9I769	Q9i769 pseudomonas
38	7	4.9	380	16	Q9WYS8	Q9wys8 thermotoga
39	7	4.9	390	5	Q95QU5	Q95qu5 caenorhabdi
40	7	4.9	396	16	O34629	O34629 bacillus su
41	7	4.9	407	17	Q9UXD3	Q9uxd3 sulfolobus
42	7	4.9	516	5	Q24282	Q24282 drosophila
43	7	4.9	518	5	Q9VUQ4	Q9vuq4 drosophila
44	7	4.9	543	3	O94342	O94342 schizosacch
45	7	4.9	753	5	Q9XZJ5	Q9xzj5 strongyloid

ALIGNMENTS

RESULT 1
Q57236
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).

```
DR   EMBL; X81714; CAA57358.1; -.
DR   EMBL; L35496; AAA23210.1; -.
DR   EMBL; X70821; CAA50152.1; -.
DR   EMBL; X99064; CAA67512.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Neurotoxin.
SQ   SEQUENCE   1278 AA;  147073 MW;  A1BE1318431D6918 CRC64;


  Query Match             100.0%;  Score 143;  DB 2;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 2e-142;
  Matches  143;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1136 NIFSNTRLYTGVEVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPE 1195

Qy     61 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1196 KIIKLIRTSNSNNSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNI 1255

Qy    121 RKNTSSNGCFWSFISKEHGWQEN 143
          |||||||||||||||||||||||
Db   1256 RKNTSSNGCFWSFISKEHGWQEN 1278


RESULT   2
Q9ZAJ5
ID   Q9ZAJ5      PRELIMINARY;      PRT;  1280 AA.
AC   Q9ZAJ5;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BONT PROTEIN.
GN   BONT.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 3281 (ATCC 43757);
RX   MEDLINE=98440323; PubMed=9767710;
RA   Santos-Buelga J., Collins M.D., East A.K.;
RT   "Characterization of the genes encoding the Botulinum neurotoxin
RT   complex in a strain of clostridium botulinum producing type B & F
RT   neurotoxins.";
RL   Curr. Microbiol. 37:312-318(1998).
DR   EMBL; Y13631; CAA73972.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1280 AA;  147487 MW;  D0F748976EBC222C CRC64;


  Query Match             15.4%;  Score 22;  DB 2;  Length 1280;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-14;
  Matches   22;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     74 SLGQIIVMDSIGNNCTMNFQNN 95
          ||||||||||||||||||||||
Db   1212 SLGQIIVMDSIGNNCTMNFQNN 1233


RESULT   3
Q45851
ID   Q45851      PRELIMINARY;      PRT;  1268 AA.
AC   Q45851;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   NEUROTOXIN TYPE F.
GN   BONT /F.
OS   Clostridium baratii.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1561;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252228; PubMed=8486245;
RA   Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA   Richardson P.T.;
RT   "Nucleotide sequence of the gene coding for Clostridium barati type F
RT   neurotoxin: Comparison with other clostridial neurotoxins.";
RL   FEMS Microbiol. Lett. 108:175-182(1993).
DR   EMBL; X68262; CAA48329.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1268 AA;  145513 MW;  963040091AC15ED2 CRC64;


  Query Match             11.2%;  Score 16;  DB 2;  Length 1268;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-08;
  Matches   16;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    127 NGCFWSFISKEHGWQE 142
          ||||||||||||||||
Db   1253 NGCFWSFISKEHGWQE 1268


RESULT   4
Q994N3
ID   Q994N3      PRELIMINARY;      PRT;    96 AA.
AC   Q994N3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   VPR PROTEIN.
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97ZA012;
RX   MEDLINE=21094715; PubMed=11177395;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
RT   "Near full-length clones and reference sequences for subtype C
RT   isolates for HIV type 1 from three different continents.";
RL   AIDS Res. Hum. Retroviruses 17:161-168(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97ZA012;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF286227; AAK30993.1; -.
```

```
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
SQ   SEQUENCE   96 AA;  11415 MW;  839CB1B0999C059B CRC64;


  Query Match             5.6%;  Score 8;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 TGVEVIIR 17
           ||||||||
Db      55 TGVEVIIR 62


RESULT   5
Q99BN5
ID   Q99BN5      PRELIMINARY;      PRT;    96 AA.
AC   Q99BN5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN.
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TV010-25;
RX   MEDLINE=21322026; PubMed=11429118;
RA   Scriba T.J., Treurnicht F.K., Zeier M., Engelbrecht S.,
RA   van Rensburg E.J.;
RT   "Characterization and phylogenetic analysis of South African HIV-1
RT   subtype C accessory genes.";
RL   AIDS Res. Hum. Retroviruses 17:775-781(2001).
DR   EMBL; AF325755; AAK09162.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
SQ   SEQUENCE   96 AA;  11450 MW;  663D5ED56DED0447 CRC64;


  Query Match             5.6%;  Score 8;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 1.5;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 TGVEVIIR 17
           ||||||||
Db      55 TGVEVIIR 62


RESULT   6
Q9BPL0
ID   Q9BPL0      PRELIMINARY;      PRT;   731 AA.
AC   Q9BPL0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FTZ-F1.
GN   FTZ-F1.
OS   Schistosoma mansoni (Blood fluke).
OC   Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC   Schistosomatoidea; Schistosomatidae; Schistosoma.
OX   NCBI_TaxID=6183;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Mendonca R.L., Bouton D., Vanacker J.-M., Laudet V., Pierce R.;
RT   "Cloning and functional characterization of a Schistosoma mansoni
RT   homologue of the FTZ-F1 nuclear receptor.";
RL   Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR   EMBL; AF158103; AAG49449.1; -.
DR   HSSP; P03372; 1HCQ.
DR   InterPro; IPR000536; Hormone_rec_lig.
DR   InterPro; IPR001723; Strdhormone_receptor.
DR   InterPro; IPR001628; zf-C4.
DR   Pfam; PF00104; hormone_rec; 1.
DR   Pfam; PF00105; zf-C4; 1.
DR   PRINTS; PR00398; STRDHORMONER.
DR   PRINTS; PR00047; STROIDFINGER.
DR   SMART; SM00430; HOLI; 1.
DR   SMART; SM00399; ZnF_C4; 1.
DR   PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW   DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW   Zinc-finger.
SQ   SEQUENCE   731 AA;  78130 MW;  20129AF9AAF30175 CRC64;


  Query Match             5.6%;  Score 8;  DB 5;  Length 731;
  Best Local Similarity   100.0%;  Pred. No. 7.6;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      68 TSNSNNSL 75
           ||||||||
Db     714 TSNSNNSL 721


RESULT   7
Q9K395
ID   Q9K395      PRELIMINARY;      PRT;  1251 AA.
AC   Q9K395;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155 (KZ 1885);
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
```

```
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 GNNCTMNF 92
           ||||||||
Db    1193 GNNCTMNF 1200


RESULT   8
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;


  Query Match             5.6%;  Score 8;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 12;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      85 GNNCTMNF 92
           ||||||||
Db    1197 GNNCTMNF 1204


RESULT   9
Q92C93
ID   Q92C93      PRELIMINARY;      PRT;    67 AA.
AC   Q92C93;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   LIN1298 PROTEIN.
GN   LIN1298.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
```

```
DR   EMBL; AL596168; CAC96529.1; -.
DR   ListiList; LIN01298; -.
KW   Complete proteome.
SQ   SEQUENCE   67 AA;  7875 MW;  E2E37AEE3BD0F3E4 CRC64;


  Query Match             4.9%;  Score 7;  DB 16;  Length 67;
  Best Local Similarity   100.0%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     39 INVVDRD 45
          |||||||
Db     31 INVVDRD 37


RESULT  10
O42077
ID   O42077      PRELIMINARY;      PRT;    96 AA.
AC   O42077;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   VPR PROTEIN.
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Song J., Wang B., Ge Y.C., Dwyer D., Dowton D., Cunnigham A.,
RA   Saksena N.;
RL   Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF000338; AAB70135.1; -.
DR   EMBL; AF000334; AAB70131.1; -.
DR   EMBL; AF000336; AAB70133.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
SQ   SEQUENCE   96 AA;  11365 MW;  99DCA046515392AF CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 GVEVIIR 17
          |||||||
Db     56 GVEVIIR 62


RESULT  11
O89653
ID   O89653      PRELIMINARY;      PRT;    96 AA.
AC   O89653;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN (FRAGMENT).
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOTHER PAIR D;
RX   MEDLINE=98325222; PubMed=9658150;
RA   Yedavalli V.R., Chappey C., Ahmad N.;
RT   "Maintenance of an intact human immunodeficiency virus type 1 vpr gene
RT   following mother-to-infant transmission.";
RL   J. Virol. 72:6937-6943(1998).
DR   EMBL; AF042966; AAC41130.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
FT   NON_TER      96      96
SQ   SEQUENCE   96 AA;  11320 MW;  5995598CDF0E7FBD CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 GVEVIIR 17
          |||||||
Db     56 GVEVIIR 62


RESULT  12
O89654
ID   O89654      PRELIMINARY;      PRT;    96 AA.
AC   O89654;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN (FRAGMENT).
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MOTHER PAIR D;
RX   MEDLINE=98325222; PubMed=9658150;
RA   Yedavalli V.R., Chappey C., Ahmad N.;
RT   "Maintenance of an intact human immunodeficiency virus type 1 vpr gene
RT   following mother-to-infant transmission.";
RL   J. Virol. 72:6937-6943(1998).
DR   EMBL; AF042967; AAC41131.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
FT   NON_TER      96      96
SQ   SEQUENCE   96 AA;  11320 MW;  5995598CDF0E7FBD CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     11 GVEVIIR 17
          |||||||
Db     56 GVEVIIR 62


RESULT  13
O89670
ID   O89670      PRELIMINARY;      PRT;    96 AA.
AC   O89670;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN (FRAGMENT).
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INFANT PAIR D;
RX   MEDLINE=98325222; PubMed=9658150;
RA   Yedavalli V.R., Chappey C., Ahmad N.;
RT   "Maintenance of an intact human immunodeficiency virus type 1 vpr gene
RT   following mother-to-infant transmission.";
RL   J. Virol. 72:6937-6943(1998).
DR   EMBL; AF042983; AAC41147.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
```

```
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
FT   NON_TER      96     96
SQ   SEQUENCE   96 AA;  11329 MW;  8DA5598CDF0E672C CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 GVEVIIR 17
           |||||||
Db      56 GVEVIIR 62


RESULT  14
O89677
ID   O89677      PRELIMINARY;      PRT;    96 AA.
AC   O89677;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN (FRAGMENT).
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INFANT PAIR D;
RX   MEDLINE=98325222; PubMed=9658150;
RA   Yedavalli V.R., Chappey C., Ahmad N.;
RT   "Maintenance of an intact human immunodeficiency virus type 1 vpr gene
RT   following mother-to-infant transmission.";
RL   J. Virol. 72:6937-6943(1998).
DR   EMBL; AF042990; AAC41154.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
FT   NON_TER      96     96
SQ   SEQUENCE   96 AA;  11320 MW;  5995598CDF0E7FBD CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 GVEVIIR 17
           |||||||
Db      56 GVEVIIR 62


RESULT  15
O89678
ID   O89678      PRELIMINARY;      PRT;    96 AA.
AC   O89678;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN (FRAGMENT).
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INFANT PAIR D;
RX   MEDLINE=98325222; PubMed=9658150;
RA   Yedavalli V.R., Chappey C., Ahmad N.;
RT   "Maintenance of an intact human immunodeficiency virus type 1 vpr gene
RT   following mother-to-infant transmission.";
RL   J. Virol. 72:6937-6943(1998).
DR   EMBL; AF042991; AAC41155.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
FT   NON_TER      96     96
SQ   SEQUENCE   96 AA;  11320 MW;  5995598CDF0E7FBD CRC64;


  Query Match             4.9%;  Score 7;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 17;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 GVEVIIR 17
           |||||||
Db      56 GVEVIIR 62


Search completed: August 15, 2002, 11:24:10
Job time: 697 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 15, 2002, 11:12:33 ; Search time 80.39 Seconds
(without alignments)
927.489 Million cell updates/sec

Title: US-08-981-087A-1
Perfect score: 431
Sequence: 1 SYTNDKILILYFNKLYKKIK.....TSSNGCFWSFISKEHGWQEN 431

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	431	100.0	1278	2	Q57236	Q57236 clostridium
2	31	7.2	1280	2	Q9ZAJ5	Q9zaj5 clostridium
3	22	5.1	1268	2	Q45851	Q45851 clostridium
4	15	3.5	1251	2	Q9K395	Q9k395 clostridium
5	15	3.5	1255	2	Q9FAR6	Q9far6 clostridium
6	11	2.6	367	2	Q45862	Q45862 clostridium
7	11	2.6	367	2	Q45861	Q45861 clostridium
8	11	2.6	1296	2	Q45894	Q45894 clostridium
9	8	1.9	96	15	Q994N3	Q994n3 human immun
10	8	1.9	96	15	Q99BN5	Q99bn5 human immun
11	8	1.9	361	2	Q45846	Q45846 clostridium
12	8	1.9	361	2	Q45848	Q45848 clostridium
13	8	1.9	441	2	Q9X708	Q9x708 clostridium
14	8	1.9	540	16	Q9A9Z8	Q9a9z8 caulobacter
15	8	1.9	731	5	Q9BPL0	Q9bpl0 schistosoma
16	8	1.9	1072	16	Q9CF64	Q9cf64 lactococcus
17	8	1.9	1291	2	Q9ZAJ8	Q9zaj8 clostridium
18	8	1.9	1291	2	Q08077	Q08077 clostridium
19	8	1.9	1291	2	Q93G71	Q93g71 clostridium
20	8	1.9	1291	2	Q933K0	Q933k0 clostridium
21	7	1.6	53	13	Q90WI9	Q90wi9 gallus gall
22	7	1.6	67	16	Q92C93	Q92c93 listeria in
23	7	1.6	96	15	O42077	O42077 human immun
24	7	1.6	96	15	O89653	O89653 human immun
25	7	1.6	96	15	O89654	O89654 human immun
26	7	1.6	96	15	O89670	O89670 human immun
27	7	1.6	96	15	O89677	O89677 human immun
28	7	1.6	96	15	O89678	O89678 human immun
29	7	1.6	96	15	O89679	O89679 human immun
30	7	1.6	96	15	Q90QQ8	Q90qq8 human immun
31	7	1.6	96	15	Q90DZ9	Q90dz9 human immun
32	7	1.6	111	13	Q90X55	Q90x55 gallus gall
33	7	1.6	131	4	Q9BUA6	Q9bua6 homo sapien
34	7	1.6	132	13	Q90X56	Q90x56 gallus gall
35	7	1.6	154	11	Q62080	Q62080 mus musculu
36	7	1.6	166	4	O95023	O95023 homo sapien
37	7	1.6	169	9	Q9G0E2	Q9g0e2 lactococcus
38	7	1.6	177	16	O51665	O51665 borrelia bu
39	7	1.6	180	15	Q9QRP6	Q9qrp6 human immun
40	7	1.6	182	15	Q85641	Q85641 moloney mur
41	7	1.6	187	15	Q85640	Q85640 murine leuk
42	7	1.6	187	15	Q83401	Q83401 moloney mur
43	7	1.6	200	16	Q9PMH8	Q9pmh8 campylobact
44	7	1.6	202	11	Q62082	Q62082 mus musculu
45	7	1.6	203	17	Q979N7	Q979n7 thermoplasm

ALIGNMENTS

RESULT 1
Q57236
ID Q57236 PRELIMINARY; PRT; 1278 AA.
AC Q57236; Q45863;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE F (BONT/F PROTEIN).
GN BONT/F.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 10281;
RA Hutson R.A., Collins M.D.;
RL Submitted (AUG-1995) to the EMBL/GenBank/DDBJ databases.
X7082
L35496
RN [2]
RP SEQUENCE FROM N.A.
RA Elmore M.J., Bodsworth N.J., Whelan S.M., Minton N.P.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 635-1000 FROM N.A.
RC STRAIN=NCTC 1028;
RX MEDLINE=94013372; PubMed=8408542;
RA Campbell K., East A.K., Collins M.D.;
RT "Gene probes for identification of the botulinal neurotoxin gene and
RT specific identification of neurotoxin types B, E, and F.";
RL J. Clin. Microbiol. 31:2255-2262(1993).
RN [4]
RP SEQUENCE OF 1-27 FROM N.A.
RC STRAIN=LANGELAND;
RX MEDLINE=98404102; PubMed=9732534;
RA East A.K., Bhandari M., Hielm S., Collins M.D.;
RT "Analysis of the botulinum neurotoxin type F gene clusters in
RT proteolytic and nonproteolytic Clostridium botulinum and Clostridium
RT barati.";
RL Curr. Microbiol. 37:262-268(1998).

DR EMBL; ~~X81714~~; CAA57358.1; -.
DR EMBL; L35496; AAA23210.1; -.
DR EMBL; X70821; CAA50152.1; -.
DR ~~EMBL; X99064~~; CAA67512.1; -.
DR HSSP; ~~P10845; 3BTR~~.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Neurotoxin.
SQ SEQUENCE 1278 AA; 147073 MW; A1BE1318431D6918 CRC64;

Query Match 100.0%; Score 431; DB 2; Length 1278;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
        1 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    848 SYTNDKILILYFNKLYKKIKDNSILDMRYENNKFIDISGYGSNISINGDVYIYSTNRNQF 907

Qy     61 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    908 GIYSSKPSEVNIAQNNDIIYNGRYQNFSISFWVRIPKYFNKVNLNNEYTIIDCIRNNNSG 967

Qy    121 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    968 WKISLNYNKIIWTLQDTAGNNQKLVFNYTQMISISDYINKWIFVTITNNRLGNSRIYING 1027

Qy    181 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1028 NLIDEKSISNLGDIHVSDNILFKIVGCNDTRYVGIRYFKVFDTELGKTEIETLYSDEPDP 1087

Qy    241 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1088 SILKDFWGNYLLYNKRYYLLNLLRTDKSITQNSNFLNINQQRGVYQKPNIFSNTRLYTGV 1147

Qy    301 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1148 EVIIRKNGSTDISNTDNFVRKNDLAYINVVDRDVEYRLYADISIAKPEKIIKLIRTSNSN 1207

Qy    361 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1208 NSLGQIIVMDSIGNNCTMNFQNNNGGNIGLLGFHSNNLVASSWYYNNIRKNTSSNGCFWS 1267

 y    421 FISKEHGWQEN 431
          ||||||||||| 
Db   1268 FISKEHGWQEN 1278
```

RESULT 2
Q9ZAJ5
ID Q9ZAJ5 PRELIMINARY; PRT; 1280 AA.
AC Q9ZAJ5;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BONT PROTEIN.
GN BONT.
OS Clostridium botulinum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 3281 (ATCC 43757);
RX MEDLINE=98440323; PubMed=9767710;
RA Santos-Buelga J., Collins M.D., East A.K.;
RT "Characterization of the genes encoding the Botulinum neurotoxin
RT complex in a strain of clostridium botulinum producing type B & F
RT neurotoxins.";
RL Curr. Microbiol. 37:312-318(1998).
DR EMBL; Y13631; CAA73972.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1280 AA; 147487 MW; D0F748976EBC222C CRC64;

Query Match 7.2%; Score 31; DB 2; Length 1280;
Best Local Similarity 100.0%; Pred. No. 1.7e-21;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    153 SISDYINKWIFVTITNNRLGNSRIYINGNLI 183
          |||||||||||||||||||||||||||||||
Db   1003 SISDYINKWIFVTITNNRLGNSRIYINGNLI 1033
```

RESULT 3
Q45851
ID Q45851 PRELIMINARY; PRT; 1268 AA.
AC Q45851;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NEUROTOXIN TYPE F.
GN BONT /F.
OS Clostridium baratii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1561;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93252228; PubMed=8486245;
RA Thompson D.E., Hutson R.A., East A.K., Allaway D., Collins M.D.,
RA Richardson P.T.;
RT "Nucleotide sequence of the gene coding for Clostridium barati type F
RT neurotoxin: Comparison with other clostridial neurotoxins.";
RL FEMS Microbiol. Lett. 108:175-182(1993).
DR EMBL; X68262; CAA48329.1; -.
DR HSSP; P10845; 3BTA.
DR MEROPS; M27.002; -.
DR InterPro; IPR000395; Bontoxilysin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF01742; Peptidase_M27; 1.
DR PRINTS; PR00760; BONTOXILYSIN.
DR ProDom; PD001963; Bontoxilysin; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 1268 AA; 145513 MW; 963040091AC15ED2 CRC64;

Query Match 5.1%; Score 22; DB 2; Length 1268;
Best Local Similarity 100.0%; Pred. No. 1.1e-12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    198 DNILFKIVGCNDTRYVGIRYFK 219
          ||||||||||||||||||||||
Db   1036 DNILFKIVGCNDTRYVGIRYFK 1057
```

RESULT 4
Q9K395
ID Q9K395 PRELIMINARY; PRT; 1251 AA.
AC Q9K395;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TYPE E BOTULINUM TOXIN.

```
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 095;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 095) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 155 (KZ 1885);
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Gyobu Y., Yamakawa K.,
RA   Kato H., Nakamura S., Karasawa T.;
RT   "C. butyricum (LCL 155) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1899;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1899) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1897;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1897) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1898;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1898) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1886;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1886) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1887;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1887) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1889;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1889) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1890;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1890) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KZ 1891;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (KZ 1891) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [11]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LCL 063;
RA   Wang X., Maegawa T., Kozaki S., Tsukamoto K., Kato H., Nakamura S.,
RA   Karasawa T.;
RT   "C. butyricum (LCL 063) gene for type E botulinum toxin.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB037714; BAB03522.1; -.
DR   EMBL; AB037704; BAB03512.1; -.
DR   EMBL; AB037705; BAB03513.1; -.
DR   EMBL; AB037706; BAB03514.1; -.
DR   EMBL; AB037707; BAB03515.1; -.
DR   EMBL; AB037708; BAB03516.1; -.
DR   EMBL; AB037709; BAB03517.1; -.
DR   EMBL; AB037710; BAB03518.1; -.
DR   EMBL; AB037711; BAB03519.1; -.
DR   EMBL; AB037712; BAB03520.1; -.
DR   EMBL; AB037713; BAB03521.1; -.
DR   HSSP; P10845; 3BTA.
DR   MEROPS; M27.002; -.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1251 AA;  143751 MW;  2021F4E427070296 CRC64;


  Query Match             3.5%;  Score 15;  DB 2;  Length 1251;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     154 ISDYINKWIFVTITN 168
           |||||||||||||||
Db     983 ISDYINKWIFVTITN 997


RESULT   5
Q9FAR6
ID   Q9FAR6      PRELIMINARY;      PRT;  1255 AA.
AC   Q9FAR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TYPE E BOTULINUM TOXIN.
GN   BONT/E.
OS   Clostridium butyricum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1492;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BL 6340/ATCC 43755/BL 5520/KZ 147;
RX   MEDLINE=20509829; PubMed=11055954;
RA   Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,
RA   Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;
RT   "Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium
RT   butyricum Strains.";
RL   Appl. Environ. Microbiol. 66:4992-4997(2000).
DR   EMBL; AB039264; BAB12249.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ   SEQUENCE   1255 AA;  143918 MW;  1B557B9D85CD8E4D CRC64;
```

```
  Query Match             3.5%;  Score 15;  DB 2;  Length 1255;
  Best Local Similarity   100.0%;  Pred. No. 8.4e-06;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    154 ISDYINKWIFVTITN 168
          ||||||||||||||
Db    986 ISDYINKWIFVTITN 1000


RESULT   6
Q45862
ID   Q45862      PRELIMINARY;      PRT;   367 AA.
AC   Q45862;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE E, HAZEN 36208 (ATCC 9564);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70815; CAA50146.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42854 MW;  0810595B3A865570 CRC64;


  Query Match             2.6%;  Score 11;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     86 NFSISFWVRIP 96
          ||||||||||||
Db    299 NFSISFWVRIP 309


RESULT   7
Q45861
ID   Q45861      PRELIMINARY;      PRT;   367 AA.
AC   Q45861;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE E (FRAGMENT).
GN   BONT/E.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE E, VH (DOLMAN);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70818; CAA50149.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     367    367
SQ   SEQUENCE   367 AA;  42902 MW;  346A610C2FF70262 CRC64;


  Query Match             2.6%;  Score 11;  DB 2;  Length 367;
  Best Local Similarity   100.0%;  Pred. No. 0.026;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     86 NFSISFWVRIP 96
          ||||||||||||
Db    299 NFSISFWVRIP 309


RESULT   8
Q45894
ID   Q45894      PRELIMINARY;      PRT;  1296 AA.
AC   Q45894; P77780;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE A (TYPE A NEUROTOXIN).
GN   BONT OR ATX.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KYOTO-F;
RX   MEDLINE=94143603; PubMed=8310180;
RA   Willems A., East A.K., Lawson P.A., Collins M.D.;
RT   "Sequence of the gene coding for the neurotoxin of Clostridium
RT   botulinum type A associated with infant botulism: comparison with
RT   other clostridial neurotoxins.";
RL   Res. Microbiol. 144:547-556(1993).
RN   [2]
RP   SEQUENCE OF 1-65 FROM N.A.
RC   STRAIN=62A;
RX   MEDLINE=97016817; PubMed=8863443;
RA   East A.K., Bhandari M., Stacey J.M., Campbell K.D., Collins M.D.;
RT   "Organization and phylogenetic interrelationships of genes encoding
RT   components of the botulinum toxin complex in proteolytic Clostridium
RT   botulinum types A, B, and F: evidence of chimeric sequences in the
RT   gene encoding the nontoxic nonhemagglutinin component.";
RL   Int. J. Syst. Bacteriol. 46:1105-1112(1996).
RN   [3]
RP   SEQUENCE OF 1-18 FROM N.A.
RC   STRAIN=TYPE A NIH;
RX   MEDLINE=96096783; PubMed=8521962;
RA   Fujita R., Fujinaga Y., Inoue K., Nakajima H., Kumon H., Oguma K.;
RT   "Molecular characterization of two forms of nontoxic-nonhemagglutinin
RT   components of Clostridium botulinum type A progenitor toxins.";
RL   FEBS Lett. 376:41-44(1995).
DR   EMBL; X73423; CAA51824.1; -.
DR   EMBL; X92973; CAA63551.1; -.
DR   EMBL; X87974; CAA61234.1; -.
DR   EMBL; D67030; BAA11051.1; -.
DR   HSSP; P10845; 3BTA.
DR   InterPro; IPR000395; Bontoxilysin.
DR   Pfam; PF01742; Peptidase_M27; 1.
DR   PRINTS; PR00760; BONTOXILYSIN.
DR   ProDom; PD001963; Bontoxilysin; 1.
KW   Neurotoxin.
SQ   SEQUENCE   1296 AA;  149410 MW;  6F12E7BF28ED69D1 CRC64;


  Query Match             2.6%;  Score 11;  DB 2;  Length 1296;
  Best Local Similarity   100.0%;  Pred. No. 0.073;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    161 WIFVTITNNRL 171
          ||||||||||||
Db   1014 WIFVTITNNRL 1024
```

```
RESULT   9
Q994N3
ID   Q994N3      PRELIMINARY;      PRT;    96 AA.
AC   Q994N3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   VPR PROTEIN.
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97ZA012;
RX   MEDLINE=21094715; PubMed=11177395;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Kalish M.L., Shaw G.M., Allen S., Hahn B.H., Gao F.;
RT   "Near full-length clones and reference sequences for subtype C
RT   isolates for HIV type 1 from three different continents.";
RL   AIDS Res. Hum. Retroviruses 17:161-168(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=97ZA012;
RA   Rodenburg C.M., Li Y., Trask S.A., Chen Y., Decker J., Robertson D.L.,
RA   Allen S., Shaw G.M., Hahn B.H., Gao F.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF286227; AAK30993.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
SQ   SEQUENCE   96 AA;  11415 MW;  839CB1B0999C059B CRC64;

  Query Match             1.9%;  Score 8;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 7.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    298 TGVEVIIR 305
          ||||||||
Db     55 TGVEVIIR 62


RESULT  10
Q99BN5
ID   Q99BN5      PRELIMINARY;      PRT;    96 AA.
AC   Q99BN5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   VPR PROTEIN.
GN   VPR.
OS   Human immunodeficiency virus type 1.
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11676;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TV010-25;
RX   MEDLINE=21322026; PubMed=11429118;
RA   Scriba T.J., Treurnicht F.K., Zeier M., Engelbrecht S.,
RA   van Rensburg E.J.;
RT   "Characterization and phylogenetic analysis of South African HIV-1
RT   subtype C accessory genes.";
RL   AIDS Res. Hum. Retroviruses 17:775-781(2001).
DR   EMBL; AF325755; AAK09162.1; -.
DR   InterPro; IPR000012; HIV_ORFXR.
DR   Pfam; PF00522; VPR; 1.
DR   PRINTS; PR00444; HIVVPRVPX.
SQ   SEQUENCE   96 AA;  11450 MW;  663D5ED56DED0447 CRC64;

  Query Match             1.9%;  Score 8;  DB 15;  Length 96;
  Best Local Similarity   100.0%;  Pred. No. 7.9;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    298 TGVEVIIR 305
          ||||||||
Db     55 TGVEVIIR 62


RESULT  11
Q45846
ID   Q45846      PRELIMINARY;      PRT;   361 AA.
AC   Q45846;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE B, NON-PROTEOLYTIC 2129B (SCOTT);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70814; CAA50145.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42175 MW;  533EA98735CD98E1 CRC64;

  Query Match             1.9%;  Score 8;  DB 2;  Length 361;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    8;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    117 NNSGWKIS 124
          ||||||||
Db    325 NNSGWKIS 332


RESULT  12
Q45848
ID   Q45848      PRELIMINARY;      PRT;   361 AA.
AC   Q45848;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
GN   BONT/B.
OS   Clostridium botulinum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TYPE B, NON-PROTEOLYTIC EKLUND 2B (COLWORTH 229);
RX   MEDLINE=94013372; PubMed=8408542;
RA   Campbell K., East A.K., Collins M.D.;
RT   "Gene probes for identification of the botulinal neurotoxin gene and
RT   specific identification of neurotoxin types B, E, and F.";
RL   J. Clin. Microbiol. 31:2255-2262(1993).
DR   EMBL; X70819; CAA50150.1; -.
DR   HSSP; P10845; 3BTA.
KW   Neurotoxin.
FT   NON_TER       1      1
FT   NON_TER     361    361
SQ   SEQUENCE   361 AA;  42131 MW;  A2E0FFFC81F9533D CRC64;
```

Query Match 1.9%; Score 8; DB 2; Length 361;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    117 NNSGWKIS 124
          ||||||||
Db    325 NNSGWKIS 332
```

RESULT 13
Q9X708
ID Q9X708 PRELIMINARY; PRT; 441 AA.
AC Q9X708;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BOTULINUM NEUROTOXIN TYPE B (FRAGMENT).
BONT/B.
Clostridium botulinum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1491;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99343691; PubMed=10413679;
RA Lalli G., Herreros J., Osborne S.L., Montecucco C., Rossetto O.,
RA Schiavo G.;
RT "Functional characterisation of tetanus and botulinum neurotoxins
RT binding domains.";
RL J. Cell Sci. 112:2715-2724(1999).
DR EMBL; AJ242628; CAB43706.1; -.
DR HSSP; P10845; 3BTA.
KW Neurotoxin.
FT NON_TER 1 1
FT NON_TER 441 441
SQ SEQUENCE 441 AA; 52772 MW; 721D0B468E8C95A4 CRC64;

Query Match 1.9%; Score 8; DB 2; Length 441;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    117 NNSGWKIS 124
          ||||||||
Db    116 NNSGWKIS 123
```

RESULT 14
Q9A9Z8
ID Q9A9Z8 PRELIMINARY; PRT; 540 AA.
AC Q9A9Z8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0813.
GN CC0813.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005758; AAK22798.1; -.
DR TIGR; CC0813; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 540 AA; 59648 MW; 72BC45442BEF99FD CRC64;

Query Match 1.9%; Score 8; DB 16; Length 540;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    238 PDPSILKD 245
          ||||||||
Db     71 PDPSILKD 78
```

RESULT 15
Q9BPL0
ID Q9BPL0 PRELIMINARY; PRT; 731 AA.
AC Q9BPL0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FTZ-F1.
GN FTZ-F1.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatoidea; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RA Mendonca R.L., Bouton D., Vanacker J.-M., Laudet V., Pierce R.;
RT "Cloning and functional characterization of a Schistosoma mansoni
RT homologue of the FTZ-F1 nuclear receptor.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; AF158103; AAG49449.1; -.
DR HSSP; P03372; 1HCQ.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Strdhormone_receptor.
DR InterPro; IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZnF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 731 AA; 78130 MW; 20129AF9AAF30175 CRC64;

Query Match 1.9%; Score 8; DB 5; Length 731;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    356 TSNSNNSL 363
          ||||||||
Db    714 TSNSNNSL 721
```

Search completed: August 15, 2002, 11:24:06
Job time: 693 sec